

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN  
COLTURE ARBOREE ED AGROSISTEMI FORESTALI ORNAMENTALI E PAESAGGISTICI  
XXV CICLO**

*SETTORE CONCORSUALE DI AFFERENZA:*

07/B2 - SCIENZE E TECNOLOGIE DEI SISTEMI ARBOREI E FORESTALI

*SETTORE SCIENTIFICO DISCIPLINARE:*

AGR/03 - ARBORICOLTURA GENERALE E COLTIVAZIONI ARBOREE

**CHARACTERIZATION OF THE GENETIC CONTROL  
OF FRUIT FLESH COLOR IN PEACH**

DISSERTAZIONE PRESENTATA DAL DOTT. **MARCO ADAMI**

*Coordinatore Dottorato*

**CHIAR.MO PROF.  
LUCA CORELLI GRAPPADELLI**

*Relatore*

**DR. STEFANO TARTARINI**

*Co-relatori*

**Dr. Alessandro Liverani  
Dr. Luca Dondini  
Dr. Paolo De Franceschi**

**Esame finale anno 2013**



# Introduction



## Peach systematics

The *Rosaceae* family includes more than 90 genera and 3000 species. Among them, in addition to ornamental species such as rose and hawthorn, are listed some of the most important fruit species like apple, pear, almond, apricot, plum, cherry, strawberry, blackberry and raspberry. Peach is classified in the order *Rosales*, family *Rosaceae*, sub-family *Prunoideae*, genus *Prunus*. Peach species (*Prunus persica* L. Batsch) together with almond (*Prunus dulcis* (Mill.) D.A.Webb) forms the subgenus *Amygdalus*, that distinguishes from other subgenera because of the presence of a deeply rough stone (Bassi & Monet, 2008). As all members of the *Prunus* genus, peach is a diploid species ( $2n=16$ ), although some haploid lines have been developed and used for breeding or research (Toyama, 1974).

*Prunus persica* is a medium sized tree, high up to 8 meters, with lanceolate, glabrous and serrated leaves. Glands are present at leaf base.

The fruit is a fleshy drupe containing a lignified, deeply-rough kernel. Unlike the almond fruit, peach mesocarp doesn't split at ripening. Being a tree species, *Prunus persica* shows a long generation time, taking 3-4 years from the seed to the first reproductive season.



Figure 1: Peach flower, fruit, seed and leaves as illustrated by A. Masclef (1891)

## The origin of modern Peach

The peach species originated in China where it has been cultivated for at least 4,000 years. The spread of peach to the west followed along ancient silk traderoutes from China to Persia (hence the name for peach, *Prunus persica*) during the second century BC (Faust & Timon, 1995). During occupation of Syria, Romans introduced peach into the Mediterranean region and subsequently it was spread from France and Italy to eastern and western Europe. Around the second half of 1500, peach was taken from Europe to south America by Spanish colonizers, and then cultivation rapidly spread in the Mexican area. By the late 1600, peach introductions are documented in Florida and, later, in both east and west coast of south America (Faust & Timon, 1995). Also, French settlers introduced peach in Louisiana, North Carolina and South Carolina. A lot of yellow and white fleshed varieties were produced commercially, leading to locally adapted populations. After the impact of Mendel's laws on the development of new breeding methodologies, north American breeders started to produce a new wave of varieties. A seedling of a new cultivar imported from China, Chinese Cling, was open pollinated by an unknown local cultivar giving origin to Elberta. As this seedling was obtained from a 'Chinese Cling' individual, with 'Early Crawford' trees in the neighborhoods, this latter cultivar has been traditionally considered the most likely pollen donor (Faust & Timon, 1995). However, SSR analysis dismantled (Aranzana *et al.*, 2010) this hypothesis, and the male parent is still unknown. When Elberta was introduced, it was an extraordinary cultivar for the commercial shipping market because of its large fruit size and superior firmness. Thereafter, it rapidly became the most popular variety in the USA, being used as part the small set of founders used by the early U.S. breeding programs (Faust & Timon, 1995).

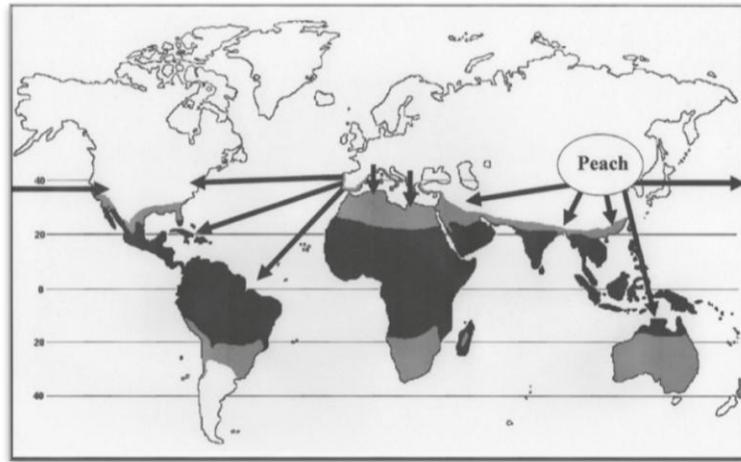


Figure 2. Early dispersion of peach (David H. Byrne et al. , 2012)

## Peach economics

World production of peaches and nectarines reached 19.4 million tons in 2012 with a growth of 3% in respect to the previous year. The top producer of peaches is China, followed by the European Union (EU) and the United States (source: USDA Foreign Agricultural Service).

With a production of 1,7 million ton/year, Italy is one of the leader countries in peach and nectarine production within the European Union. More than 90000 ha are dedicated to peach growth, with a relatively standard trend within the last five years, with 35% dedicated to nectarine and 65% to peaches (source: ISTAT <http://agri.istat.it>). Emilia Romagna and Campania are the regions where the production is most concentrated. Emilia Romagna is also the region where the cultivation of nectarines is growing at expenses of peaches, following a trend already seen in California, where in recent years the production of peaches is steadily declining in favor of nectarines (Pirazzoli, 2008).

In terms of total production, there are no big differences between Northern and Southern Italy. In the North areas are mainly grown medium ripening varieties, whose productivity greatly exceeds that of the early cultivars grown in the South, where the production is mainly destined to the fresh consumption (Pirazzoli, 2008).

## Mutations and chimerism

Chimeras arise when a cell located near the crest of the apical meristem in a bud, undergoes mutation. All the cells which are later produced by mitosis from it will carry this mutation.

The apex is organized into a layered region (the tunica) and a region where layering is not evident (the corpus). The cells are arranged in three main different and independent layers that originate all the tissues of the bud and eventually to all organs of the plant.

The controlled pattern of cell divisions in the tunica results in the maintenance of discrete layers, with the number of layers varying somewhat among the different species (Szymkowiak & Sussex, 1996).



Figure 3. Layered organization of vegetative apex (adapted from Dermen, 1960)

The derivatives of the outermost layer I (LI) give rise to the epidermis, a continuous layer that cover all tissues of the leaf, stem, flower petals, etc. Derivatives of layer II (LII) give rise to several layers within the stem, a large proportion of the cells in the leaf blade, reproductive organs and gametes. Derivatives of layer III (LIII) give rise to most of the internal tissue of the stem and a number of cells around the veins within the leaf. Layer LII and LIII produce cells both by anticlinal and periclinal mitosis, while LI only shows anticlinal division. Moreover, cells originating from different layers are distinguished not only by their division plan, but also by size, vacuolization and proliferative speed (Szymkowiak & Sussex, 1996).

Chimeral plants can be categorized on the basis of the location and relative proportion of mutated to non-mutated cells in the apical meristem.

### *Periclinal chimeras*

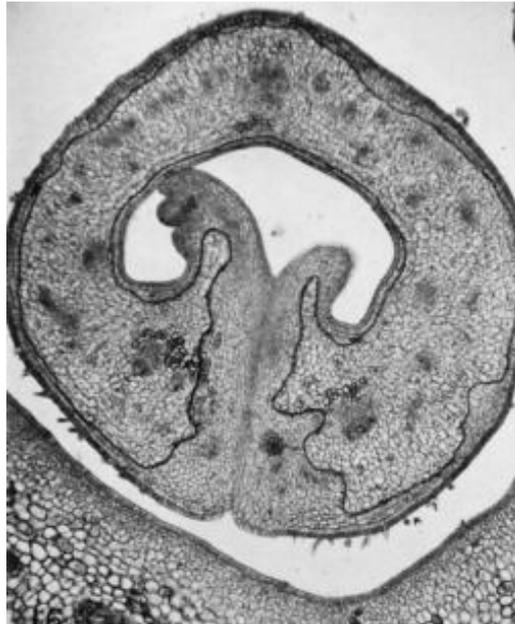
Periclinal chimeras are the most important category since they are relatively stable and can be maintained by vegetative propagation; moreover if the mutated layer is LII, they can even be sexually transmitted. A mutation produces a periclinal chimera if the affected cell is positioned near the apical dome so that the cells produced by subsequent divisions form an entire layer of the mutated type (Szymkowiak & Sussex, 1996). The resulting meristem contains one layer which is genetically different from the remainder of the meristem. If, for example, the mutation occurs in LI, then the epidermal layer of the shoot which is produced after the mutation is the new genetic type.

### *Mericlinal chimeras*

Mericlinal chimeras are produced when the derivatives of the mutated cell do not entirely cover the apical dome. A mutated cell layer may be maintained on only one portion of the meristem giving rise to chimeral shoots or leaves which develop in that portion, while those that differentiate on all other portions of the meristem are normal, nonchimeral shoots. Many mericlinal chimeras involve such a limited number of cells that only a small portion of one leaf may be affected. Like periclinal chimeras, mericlinal chimeras are generally restricted to one cell layer.

### *Sectorial chimeras*

Sectorial chimeras result from mutations which affect sections of the apical meristem, the altered genotype extending through all the cell layers. This chimeral type is unstable and can give rise to shoots and leaves which are not chimeral. Both normal and mutated solid shoots can be produced, depending upon the point on the apex from which they differentiate.



*Figure 4. Ovary of a chimerical peach: the region inside the ink line (LIII) is tetraploid (Dermen & Stewart, 1973)*

*Fruit flesh chimeras in peach*

Several mutations of the flesh color are reported in literature, both from white to yellow (i.e. Springwhite/Springcrest and Caldesi 2000/Cristina) and viceversa (i.e. Maycrest/Whitecrest, Armking/Silverking and Redhaven/White Redhaven). The mutations in White Redhaven and Cristina are periclinal as in the fruit of these cultivar monolayer epidermis folds down and invaginates inside suture giving origin to a thin LI sector within the LII originated flesh (Figure 4; Dermen & Stewart, 1973) that maintain the ancestral color (Figure 5).



*Figure 5. Evidences of the periclinal origin of White Redhaven (A) and Cristina (B) fruits. The ancestral color is still visible in a thin sector of the flesh in correspondence of the fruit suture.*

## Phenotypic traits

While initial breeding efforts were aimed to improve aesthetic and technological traits (color, firmness and attractiveness), modern peach breeders have focused on tree productivity, fruit quality and reduction of production costs by improving agronomic and disease resistance traits (Bassi & Monet, 2008). In a number of cases, many traits have been genetically characterized as being “mendelian” (also “single” or “qualitative”), i.e. attributable to the action of two alleles at one locus. Some examples of genes with a mendelian inheritance in peach are reported in. However, most of the fruit quality traits are controlled by more than one gene, with a continuous (not discrete) distribution in segregating progenies and the trait variability is also influenced by environmental conditions. The genomic regions containing genes that are associated with a determined quantitative character are called “Quantitative Trait Loci” (QTL).

Peach trees are used both for fruit production and ornamental purpose. Different typologies of tree are classified on internode length and tree growth.

Standard internode length varies between 15 and 25 cm; dwarf phenotypes, showing a reduced internode size, depend on several loci, most of them monogenic (Gradziel & Beres, 1993; Bassi & Monet, 2008; Ogundiwin *et al.*, 2009). Tree growth habit is related both to mendelian and polygenic traits. The most common tree form are standard, arching, columnar, compact, open, spreading, spur, twister, upright and weeping.

Narrow leaf phenotypes are found within germplasm, and are usually linked to dwarf genes that are associated to mendelian loci (Chaparro *et al.*, 1994). The leaf blade can be flat or wavy, the latter determined by a recessive allele at the *Wa* locus (Scott & Cullinan, 1942).

The form of glands at leaf base is controlled by the mendelian *E* locus with incomplete dominance. Glands are reniform when the dominant allele is homozygous while, in heterozygotes, the form is globose and glands are absent in recessive homozygotes. Noteworthy is the fact that the presence of the dominant allele is correlated to the tolerance to powdery mildew. Leaf color phenotype of red colored cultivars depends by the expression of different monogenic loci. The *anthocyanin deficiency* (*An/an*) and *anthocyaninless* (*W/w*) genes also affect the color of leaf, flower and fruit (Bassi & Monet, 2008).

Finally, senescent leaves color is linked to fruit flesh color: white fleshed genotypes exhibit yellow senescent leaves, while yellow-fleshed types show orange senescent leaves (Williamson *et al.*, 2006).

Peach has hermaphroditic, perigynous flowers with gamosepalous calyx that spontaneously falls during fruit development (Bassi & Monet, 2008). The species is usually autofertile with entomophilous pollination, even if a few male-sterile cultivars exist. This trait is conditioned by two recessive loci: *ps* (homozygous in 'J. H. Hale'; Bailey & French, 1949) and *ps2* (found in cv. 'White Glory'; Werner & Creller, 1997).

Petal color can vary from pure white to dark red and variegated in red leafed varieties. Also in some ornamental cultivars chrysanthemum-like petal has been described (Yoshida *et al.*, 2000). Flower can show differences in petal numbers. Normal flowers have typically 5 petals, while in semi-double flowers some stamens are transformed to petals resulting in flowers with 12-24 petals; in double flower phenotypes also sepals are transformed in petals, in addition to stamens. Two types of corolla can be found: "showy" (rose shaped, large petals) and "non-showy" (bell shaped, small petals). The inheritance of this trait is monogenic, being non-showy the dominant (*Sh/sh*). Within the "showy" phenotype a second gene (*L/l*) controls the size of petals, being the large-sized showy trait the dominant (Connors, 1920).

The color of the hypanthium (the inner part of the calyx) is controlled by the same locus (*Y*) that controls fruit flesh color. White fleshed individuals have yellow hypanthia, while yellow-fleshed types show orange hypanthia (Williamson *et al.*, 2006).

### *Fruit traits*

Fruit weight usually varies from 180 to 230 g at harvest time in commercial varieties. It is an important quantitative trait that shows a significant genetic component (Etienne *et al.*, 2002; Quilot & Kervella, 2005), being also affected by the pleiotropic action of some major genes (Eduardo *et al.*, 2010).

Together with the common round-shaped peaches, are becoming popular "flat" peaches, whose fruit appears flattened at opposite poles. The saucer trait is monogenic dominant (*S*) over round peach (*s*), and the homozygous is lethal (Lesley, 1940). The flat genotype, in addition to seed shape, also influences negatively germination rate and viability.

Phenotype and symbol	Genotype	Linkage group		Note	Reference
<b>Fruit traits</b>					
Slow ripening (Sr)	sr/sr	-	-		Ramming (1991)
Saucer or flat shape (S)	S/-	6	S/S is lethal		Lesley (1940)
Aborting fruit (Af)	af/af	6	-		Dirlewanger et al. (2006)
Blood red fruit (Bf)	Bf/-	-	-	Pigment appears in immature fruit and main leaf vein; often smaller trees	Werner et al. (1998)
Rough skin (Rs)	rs/rs	-	-	Matte skin surface; glabrous flower buds	Okie and Prince (1982); Okie (1988b)
Nectarine (Glabrous skin, G)	g/g	5	Fuzzless		Blake (1932); Blake and Connors (1936)
Full red skin (Fr)	fr/fr	-	-	Expressed just in fruit	Beckman and Sherman (2003)
Highlighter (H)	h/h	-	-	Red colour suppression on fruit skin	Beckman et al. (2005)
White flesh (Y)	Y/-	1	-	Also affects hypanthium and leaf colour	Connors (1920)
<b>Flesh texture and pit adherence (F)</b>					
Melting freestone	F/-	4	-		Bailey and French (1932, 1949); Monet (1989); Peace et al. (2005)
Melting clingstone	f/f	-	-		Peace et al. (2005)
	f/f1	-	-		
	f/n	-	-		
Non-melting clingstone	f1/f1	-	-		Peace et al. (2005)
	f1/n	-	-		
	n/n	-	-		
Stony hard flesh (Hd)	hd/hd	-	-		Yoshida (1976); Scorza and Sherman (1996)
	hdhd/F	-	-	Stony hard, melting	Haji et al. (2005)
	hdhd/f1f1	-	-	Stony hard, non-melting	Haji et al. (2005)
Low-acid flesh (D)	D/-	7	-	D for "douce" (sweet in French)	Monet (1979)

Table 1 Mendelian traits related with fruit discovered in peach. Modified from Bassi & Monet, 2008)

Melting (M) and non-melting flesh (NM) are the most known fruit flesh textures in peach. Melting flesh peaches mesocarp undergo a strong softening in the last stage of ripening, while non-melting maintain a firm texture until full ripening and slowly soften towards senescence. This trait is affected by the cell-wall composition and metabolism. The difference between the two flesh types depends on a lack of endopolygalacturonase (endoPG) activity, one of the enzymes responsible for cell wall disruption during the softening process, although the typical climacteric increase of ethylene is present in both flesh types (Mignani *et al.*, 2005). This trait results to be highly associated to the freestone/clingstone trait (flesh adherence to the pit), another commercially important criterion to classify peach cultivars (Peace & Ahmad, 2004; Morgutti *et al.*, 2006).

Fruit epidermis is pubescent in standard peaches, while glabrous in nectarines. The nectarine phenotype is monogenic recessive (*g*) and probably had origin in the north west of China (Faust & Timon, 1995). The smooth skin makes nectarine fruit more susceptible to mechanical bruising and pest damage.

Epidermis color is determined by two main pigments, carotenes and xanthophylls, that give the orange and yellow ground color, and the anthocyanins, responsible of red/blue over-color. The red over-color is a quantitative character, influenced by light exposure and ripening. Two loci that affect the fruit red color have been identified: the "redleaf"

gene, expressed also in leaves and *Rf/rf*, expressed in fruit skin only (Beckman & Sherman, 2003). A third mendelian locus controlling the trait is “highlighter” (H/h), whose recessive allele suppresses the presence of skin over-color (Beckman & Alcazar, 2005).

### *Peach flesh color*

Flesh color is one of the most commercially important traits in peach fruits. Cultivars are in fact classified into two main groups: white and yellow peaches. A third flesh color phenotype is present within peach germplasm: “Red blood-flesh”. These varieties are characterized by a red stain in almost all the flesh, independent of the ground color: this trait has been described as dominant (Werner *et al.*, 1998). White peaches show reduced or absent carotenes and xanthophylls content and have a distinct flavour compared to yellow peaches. However yellow peaches are often preferred by consumers, possibly because of their higher concentration in orange carotenoids that could mask flesh oxidation caused by blemishes (Pirazzoli. C, 2008). The intensity of yellow color in the mesocarp is very variable among peach germplasm. It is in fact known that this trait correlates with the carotenoid (in particular the  $\beta$ -carotene) content of the flesh. Yellow fleshed varieties show a  $\beta$ -carotene content of 2-3 mg/100g of fresh weight, whereas white fleshed varieties show a reduction in carotenoid up to 10 fold, ranging from 0,01 mg to 1,8 mg (Vizzotto *et al.*, 2006).

Studies on yellow flesh color segregation showed that this is a simple mendelian trait controlled by a single locus (Y) where the recessive allele brings the yellow flesh and the dominant allele the white flesh (Connors, 1920). The latter is believed to be the ancestral one, that subsequently gave origin to the yellow flesh phenotype. This hypothesis is supported by the fact that in the species origin area there is a high predominance of white fleshed genotypes (Faust & Timon, 1995). In 1990 Morrison highlighted the tight correlation between  $\beta$ -carotene content in fruit and leaves, allowing allows the early selection of seedlings using the leaf color as reference. This link allowed the identification of molecular markers associated to the leaf color trait (Warburton *et al.*, 1996). The Y locus was preliminary mapped on a few plants of an existing cross population, Royal Prince  $\times$  Yoshihime (RP  $\times$  Yo), that segregates for flesh color (Brandi PhD thesis, 2010). The position of Y locus was known to be located on the peach linkage group 1 (Aranzana *et al.*, 2003). The Brandi’s thesis work allowed to map

(Figure 6) the position of *Y* locus between two SSR markers, *pchgms3* (Sosinski *et al.*, 2000) at 9 cM from the locus and *PacA18* (Decroocq *et al.*, 2003) at 11,7 cM.

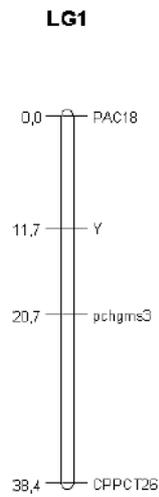


Figure 6. Frame of *RP x Yo* map showing the location of *Y* locus (Brandi PhD thesis)

The *Y* locus was also independently mapped on LG1 (Cao *et al.*, 2010) in linkage group 1 at about 7 cM from the SSR marker CPDCT024, that is close to the *pchgms3* used in Brandi (2010).

## Carotenoids

Carotenoids are a class of of terpenoid pigments that can be mainly found in plants and photosynthetic organisms. Like all terpenes, they originate by the union of isoprenic units, that are synthesized from acetyl-CoA or glycolysis pathway intermediates(Cunningham & Gantt, 1998). The number of 5-carbon units combined gives the classification of terpenes as summarized in Table 2.

Terpene	N. of isoprene unit	N. of C atoms
<i>Monoterpenes</i>	2	10
<i>Sesquiterpenes</i>	3	15
<i>Diterpenes</i>	4	20
<i>Sesterterpenes</i>	5	25
<i>Triterpenes</i>	6	30
<i>Sesquarterpenes</i>	7	35
<i>Tetraterpenes</i>	8	40
<i>Polyterpenes</i>	>10	>50

Table 2: Composition of terpenoids (Cunningham & Gantt, 1998)

Carotenoids are tetraterpenes that contain 40 carbons arranged in a polyene hydrocarbon chain. They are normally divided in oxygen-containing xanthophylls and carotenoids (Lu & Li, 2008). The latter can be linear or cyclic, depending on the presence of carbon rings ( $\beta$  or  $\epsilon$ ) at molecule edge (i.e.  $\beta$ -carotene). Colored carotenoids adsorb wavelengths between 400nm and 500nm, therefore their color ranges from pale yellow through bright orange to deep red, and is directly linked to their structure. The presence on conjugated double-bonds in fact, allows electrons to move freely across these areas of the molecule. As the number of conjugated double bonds increases, the energy needed by the electrons to change states decreases. This causes the range of energies of light absorbed by the molecule to decrease as well. The more frequencies of light are absorbed from the short end of the visible spectrum, the more the compound acquires an increasingly red appearance. The pigmentation appears when the molecule has are at least seven conjugated bonds (Lu & Li, 2008).

More than 600 different structures are found in nature and are responsible for coloration of many organisms. In plants carotenoids are found within thylakoid

membranes in chloroplasts and in lipid bodies within chromoplasts, the organelles responsible for flowers and fruit pigmentation (Walter & Strack, 2011).

#### *Main roles of carotenoids*

*Photoreception* - In photosynthetic organisms, carotenoids play a central role in the photosynthesis process. In plants,  $\beta$ -carotene acts as accessory pigment in light harvesting complexes (LHC, LHC II). The polyene structure makes it capable of absorbing a broader range of wavelengths in the blue region of the visible spectrum than chlorophyll and then transfer the energy to chlorophyll (Giuliano *et al.*, 1993).

*Photo-protection* – Beyond the participation in the energy-transfer process, carotenoids play a vital role in the photosynthetic reaction center. They provide a mechanism for photo-protection against auto-oxidation by quenching triplet state chlorophyll molecules and scavenging singlet oxygen and other toxic oxygen species formed within the chloroplast during photosynthesis (Auldridge *et al.*, 2006). The reason for this photo-protective ability resides in the high number of conjugated double-bond, that make them capable of dissipating the excess of energy as heat. Zeaxanthin and xanthophylls (Niyogi *et al.*, 1997) take part to the xanthophyll cycle, a mechanism that reduce the amount of energy that reaches the photosynthetic reaction centers, hence protecting the photosynthetic tissues against photo-oxidative damages. During light stress violaxanthin is converted to zeaxanthin via the intermediate antheraxanthin. This compound plays a direct photo-protective role by acting as a lipid-protective anti-oxidant and by stimulating non-photochemical quenching within light-harvesting proteins. This conversion of violaxanthin to zeaxanthin is done by the enzyme violaxanthin de-epoxidase, while the reverse reaction is performed by zeaxanthin epoxidase (Niyogi *et al.*, 1997).

*Abscissic acid biosynthesis* - Abscisic acid (ABA) is a plant hormone that plays an important role in the regulation of drought tolerance, seed development and sugar sensing. The backbone of this compound is a cleavage product of 9-cis-epoxycarotenoids by a 9-cis-epoxycarotenoid dioxygenase (NCED). This initial cleavage step that leads to xanthoxin, is also the limiting step where the biosynthesis of ABA is mainly regulated (Auldridge *et al.*, 2006).

## Carotenoid biosynthesis pathway

The core carotenoid pathway is conserved in most plant species although some plants accumulate special and rare carotenoids via unique biosynthetic routes. As isoprenoids, carotenoid compounds originate in the plastid-localized 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway that starts with the reaction between pyruvate and glyceraldehyde-3-phosphate (Farré *et al.*, 2010). The first steps in the MEP pathway are regulated by 1-deoxy-D-xylulose-5-phosphate synthase (DXS) and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) (Fig. 7).

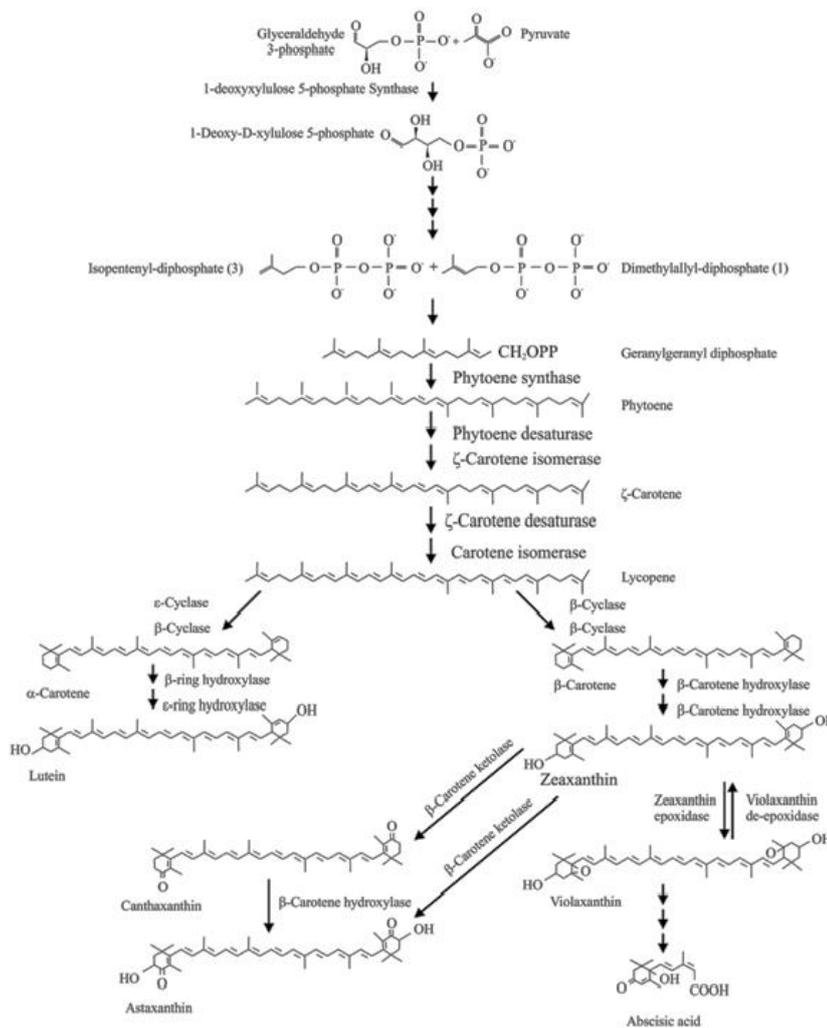


Fig. 7: Carotenoid biosynthesis pathways (Cunningham & Gantt, 1998)

These first reactions are a first regulatory step in carotenoid biosynthesis (Farré *et al.*, 2010). Overexpression of DXS in *Arabidopsis* seedlings was shown to result in up to 112–131% increase in the total carotenoid content, whereas silencing of this gene reduced

the carotenoid content by 75–87% relative to the wild type control. In ripening tomato fruits, the abundance of *dxs* transcript was found to be developmentally regulated, and corresponded to changes in expression of *psy* and accumulation of carotenoids (Lu & Li, 2008).

The second key regulatory step is catalyzed by 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (HDR) eventually leading to the production of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (Fig. 7). Geranyl-geranyl diphosphate (GGPP) synthase catalyses the condensation of three molecules of IPP and one molecule of DMAPP to produce GGPP - a 20-carbon molecule (Farré *et al.*, 2010). Subsequently (Fig. 7), two molecules of GGPP are condensed by phytoene synthase (PSY) to form phytoene (Farré *et al.*, 2010).

Phytoene biosynthesis is considered one of the main regulatory step in the pathway, a real “bottleneck”. Most of plant species express multiple functionally redundant copies of phytoene synthase (*psy*), although different *psy* genes appear to be differentially expressed and regulated (Farré *et al.*, 2010). For example, a phytochrome-interacting transcription factor, RIF, binds to the *psy* promoter and maintains it in a repressed state under dark conditions. Under light conditions, RIF degrades and dissociates from the *psy* promoter, thus allowing its active expression (Toledo-Ortiz *et al.*, 2010). Overexpression of the maize *psy* gene resulted in higher carotenoid levels in the rice endosperm (Paine *et al.*, 2005). Phytoene then undergoes four sequential reactions to form lycopene, catalyzed by phytoene desaturase (PDS). At this step, the carotenoid biosynthesis pathway branches (Cunningham & Gantt, 1998). One branch forms carotenoids with two  $\beta$ -rings, while the other introduces both  $\beta$ - and  $\epsilon$ - rings to lycopene to form  $\alpha$ -carotene, which is then converted to lutein (Cunningham & Gantt, 2001). The relative activities of  $\beta$ -CYC and  $\epsilon$ -CYC determine the proportion of lycopene channeled to the two branches of the carotenoid pathway and act as a major regulatory step in carotenoid biosynthesis by determining the ratios of  $\beta,\beta$ - and  $\epsilon,\beta$ -carotenoids. (Cunningham *et al.*, 1996; Pogson *et al.*, 1996; Cazzonelli *et al.*, 2010).

### *Carotenoid accumulation*

The differential accumulation of carotenoids in plant tissues may depend on the three distinct processes of biosynthesis, compartmentalization and degradation (for a review, see Walter & Strack, 2011). In tomato (*Solanum lycopersicum*) fruit, marigold (*Tagetes erecta*) flower and canola (*Brassica napus*) seed, the white phenotype arose as a

consequence of the lower expression of phytoene synthase, the enzyme catalyzing the first committed and rate-limiting step in carotenoid biosynthesis (Fray & Grierson, 1993; Shewmaker *et al.*, 1999; Moehs *et al.*, 2001). In cauliflower (*Brassica oleracea var. Botrytis*), the insertion of the *Or* gene (encoding a plastid-targeted protein containing a cysteine-rich zinc finger domain) triggered the differentiation of uncolored plastids into carotenoid-containing chromoplasts, changing tissue color from white to orange (Zhou *et al.*, 2008).

Finally, carotenoid accumulation is influenced by the oxidative cleavage activity of degradative enzymes. This process produces an array of terpenoid products collectively known as apocarotenoids. These include abscisic acid and strigolactones, and other volatile and non-volatile compounds, which are well known for their use as aromas, flavors and fragrances. Some apocarotenoids, e.g.  $\beta$ -ionone, are also known to play a role in plant-insect interactions.

Carotenoid degradation is catalyzed by three main classes of dioxygenases involved in several biosynthetic processes (Figure 8). The CCD7/CCD8 are a class of dioxygenases involved in the generation of the apocarotenoid hormone strigolactone (Walter & Strack, 2011).

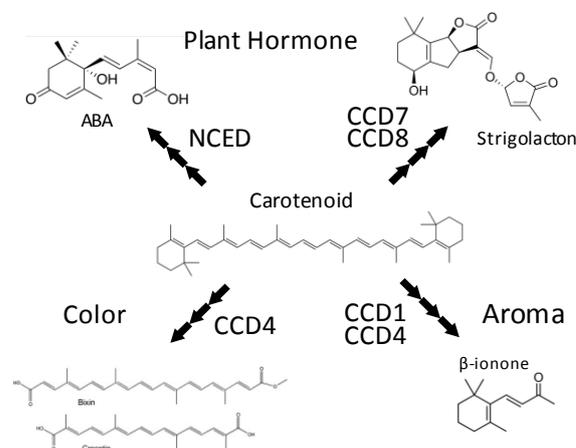


Figure 8. Apocarotenoids and the dioxygenases responsible for their production. Adapted from (Auldridge *et al.*, 2006)

The ABA-related NCEDs (9-*cis*-epoxycarotenoid dioxygenases) and the carotenoid cleavage dioxygenases, CCD4 and CCD1, symmetrically cleave double bonds at 9,10 (9',10') positions of the substrates and are known to be involved in volatile compounds release (Walter & Strack, 2011). CCD1s are cytosolic enzymes that seem to act on a broader range of substrates, while most CCD4 enzymes are targeted to plastids

(Ytterberg *et al.*, 2006; Rubio *et al.*, 2008; Walter & Strack, 2011). In fleshy fruits, carotenoids are generally accumulated in plastoglobules within chromoplasts (Bian *et al.*, 2011; Klee & Giovannoni, 2011); thus, plastid targeting of CCD4 enzymes provides access to the main site of carotenoid accumulation in flowers and fruits. In chrysanthemum (*Chrysanthemum × morifolium*), petals of white and yellow varieties did not show differences in the expression of carotenogenic genes, but carotenoids are degraded by a petal-specific CCD4 whose gene is absent in yellow varieties (Ohmiya *et al.*, 2006; Yoshioka *et al.*, 2011). In potato (*Solanum tuberosum*) the expression of a *ccd4* was reported to be higher in white-fleshed tubers than in yellow-fleshed ones (Diretto *et al.*, 2007). Similarly, in saffron (*Crocus sativus*) the expression of a *ccd4* gene was high in the white portion of the stigma and very low in the orange part, rich in carotenoids; also, CCD1 enzymes resulted less active than CCD4s in this system (Rubio *et al.*, 2008; Gómez-Gómez *et al.*, 2010). Therefore, in the above mentioned cases in which a color change occurred as a consequence of carotenoid degradation, it seems to be controlled by a CCD4 rather than CCD1 or NCED enzymes.

#### *Differential carotenoid accumulation and gene expression in peach*

Brandi *et al.*, 2011 made a comparison of the cultivar Redhaven and its white-fleshed mutant White Redhaven in terms of carotenoid content and related genes expression. It has been observed that the carotenoid content of the two accessions differs since the S3 stage of ripening and it is maximum at full ripening, when the yellow Redhaven contains ten times more carotenoids (mainly  $\beta$ -ring type) than its white sport. This strong difference between the two clones is reflected also in the expression pattern of the carotenoid-related genes analyzed; in the White Redhaven (Figure 9) the expression of the early pre-pathway genes is generally low during maturation, with the exception of 1-deoxy-d-xylulose 5-phosphate synthase (DXS) that peaks at stage S2 and remains constantly high, while the early pathway genes show a constant increase until maturity; Brandi *et al.*, 2011 (Figure 9) suggested that the up-regulation of these genes is due to a feedback regulation dependent on the low carotenoid level or their degradation products. Late pathway genes show a constant low expression, with the exception of carotene  $\beta$ -hydroxylase (*chy-b*), that is strongly up-regulated. Among the carotenoid cleaving dioxygenases analyzed, only the carotenoid cleavage dioxygenase 4 (*ccd4*) has an high level of expression since stage S3 (Figure 9). The up-regulation of *chy-b* and specially *ccd4* are negatively linked with the accumulation of  $\beta$ -ring carotenoid (Diretto

*et al.*, 2007; Yan *et al.*, 2010). In Redhaven the expression of the early pre-pathway genes is generally low during the maturation, similarly to White Redhaven. However, the *dxs* after the S2 stage peak, drops back to the previous levels until maturity (Figure 9). Similarly, the early pathway genes *psy* and *zds* show a peak in S3 and then fall back, whereas in White Redhaven the increase rate is stable until maturity.

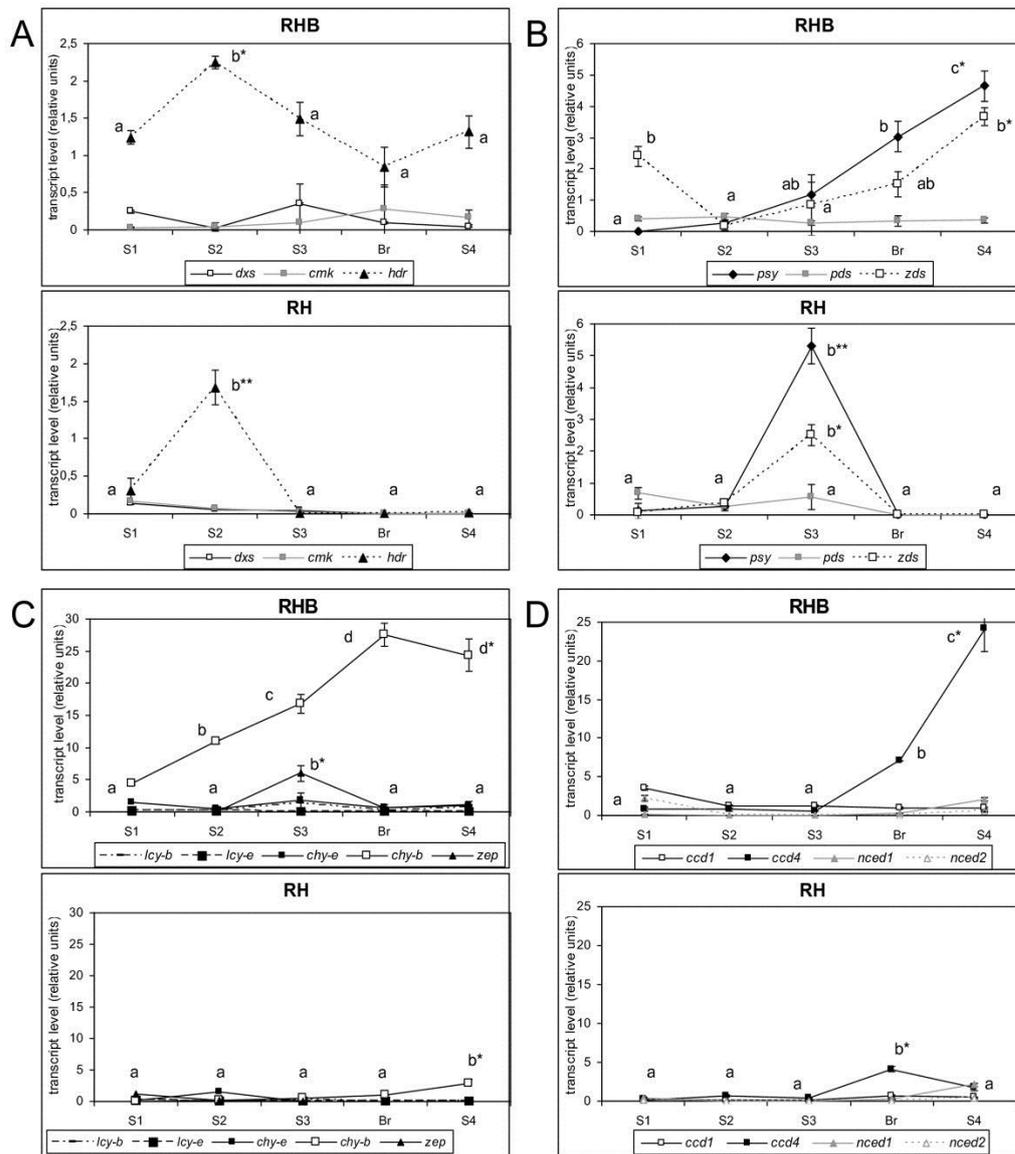


Figure 9. Differential expression of carotenoids pathway genes during ripening of RHB and RH fruits. Relative average gene transcript levels normalized with *rps28* values. A: isoprenoid genes [*cmk*, 4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol kinase; *dxs*, 1-deoxy-d-xylulose 5-phosphate synthase; *hdr*, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase]. B: early carotenoid genes (*pds*, phytoenadesaturase; *psy*, phytoene synthase; *zds*,  $\zeta$ -carotene desaturase). C: other carotenoid genes (*chy-b*, carotene b-hydroxylase; *chy-e*, carotene  $\epsilon$ -hydroxylase; *lcy-b*, lycopene b-cyclase; *lcy-e*, lycopene-e-cyclase; *zep*, zeaxanthin epoxidase). D: dioxygenase-related genes (*ccd1* and *ccd4*, carotenoid cleavage dioxygenases 1 and 4; *nced1* and *nced2*, 9-cis-epoxycarotenoid dioxygenases 1 and 2). For each gene, different letters indicate significant differences among mean values from different stages (\*:  $p \leq 0.05$ ; \*\*:  $p \leq 0.01$ ) (From Brandi *et al.*, 2011).

## Molecular markers and mapping

Genetic markers are readily assayed phenotypes that have a direct correspondence with DNA sequence variation at a specific location in the genome or locus. The assay for a genetic marker is not affected by environmental factors but one marker can be differentially adoptable depending on the phenological status.

For mapping, the ideal genetic marker is codominant (allow discrimination of homozygous and heterozygous allelic configurations), multiallelic, and highly polymorphic within the species.

Advances in molecular techniques allowed the direct detection of DNA sequence variation and led to the development of a vast range of molecular markers that have largely replaced morphological and biochemical markers. Their detection is mostly based on electrophoresis or hybridization techniques. The great advantages of molecular markers over morphological and biochemical are the ubiquity and uniformity of diffusion all along the genome and the independence from environmental or phenological stage. Some molecular markers, RFLP (Restricted Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA) and AFLP (Amplified Fragment Length Polymorphism) allow to simultaneously assay numerous loci, while SSR are usually able to distinguish many different alleles at the same locus. SNP (Single Nucleotide Polymorphism) detect polymorphism at single nucleotide level.

### *SSR markers*

SSR markers, also called “microsatellites” (Litt & Luty, 1989), correspond to tandem repetitions of di-, tri-, tetra-, penta-, hexa- and hepta-nucleotides. The number of repetitions is variable between alleles, which can be distinguished based on their size after PCR amplification, yielding co-dominant markers.

After their discovery in humans (Hamada, 1982), SSR were identified in eukaryotes and prokaryotes where they appear to be ubiquitous and frequent across the genome, especially in non-coding transcribed regions (Morgante *et al.*, 2002). While genotyping is relatively straightforward (through a simple PCR amplification, electrophoresis band resolution and detection), this technique requires the previous development of the primers flanking the repetitive region. This can be obtained by sequencing of SSR-enriched genomic or cDNA libraries or genomic/transcriptomic sequence information.

Nearly 600 SSR primer sequences are available in the *Prunus* genus (available at the Genome Database for Rosaceae, [www.rosaceae.org](http://www.rosaceae.org)). Furthermore, they have been extensively used in genetics and evolution studies of various species (reviewed in Agarwal *et al.*, 2008; Pleines *et al.*, 2009).

Although cases of detection of more than one locus (Dirlewanger *et al.*, 2002) have been reported, in most cases one PCR analysis yields just one marker, implying that the cost of SSR markers is high compared to other techniques that generate information for multiple loci in one assay (e.g. AFLP).

Novel high-throughput techniques that use capillary electrophoresis devices, such as the ABI Prism 3730 DNA Analyzer, allow multiplexing by using four different fluorophores, reducing significantly the cost of the assay. This is obtained by the simultaneous amplification of up to six markers per colour instead of using just one labelled primer per marker in the case of classical SSR analysis. Finally, these markers are extremely robust and highly reproducible, being these key conditions for linkage mapping (Kumar *et al.*, 2009).

In 2010 an extensive analysis on 212 peach commercial varieties was carried out to study the genetic variation of European and American peach germplasm (Aranzana *et al.*, 2010). The analysis showed how genotypes stratificate in three main subpopulations, approximately corresponding to the fruit differences: Melting flesh peaches, Melting flesh nectarines and Non Melting varieties. It showed furthermore how the linkage disequilibrium (LD) in these peach germplasm is high, up to 13–15 cM. This depends both on the peach self-compatibility and on the small set of US founders used by the early US breeding programs (Aranzana *et al.*, 2010).

## SNP

Single nucleotide sequence variations (substitutions, insertions and deletions) provide an almost unlimited source of co-dominant markers. They exhibit a high variability and abundance across plant genomes (Varshney *et al.*, 2009). Even if they have been developed more recently than the other types of markers, they have been quickly adopted and applied for a wide range of objectives, e.g., linkage mapping (Chagné *et al.*, 2008), association genetics (Chu *et al.*, 2009) and genome evolution (Garvin *et al.*, 2010). SNP discovery is efficiently performed through whole genome/transcriptome re-sequencing of different cultivars (Varshney *et al.*, 2009), either from sequencing specifically targeted loci (e.g., candidate genes or ESTs; (Rafalski, 2002).

A first approach for SNP discovery was sequencing on capillary electrophoresis, but at present the usually called “next generation sequencing” (NGS, reviewed in Schuster, 2008) represent the most powerful approach for SNP discovery. By sequencing about 600 Gbp per run it is possible to identify SNPs not only in gene coding and non-coding sequences, but also in intergenic regions.

If it falls in a coding region, the SNP can generate a synonymous mutation or a non-synonymous mutation that results in an aminoacid exchange (missense mutation) or in a stop codon (nonsense mutation). Therefore, besides serving as genetic markers, SNPs may be directly associated with phenotypic variation thus providing functional markers. NGS technologies open the possibility to detect markers in the whole genome or transcriptome with a high density with an affordable price if entire populations have to be analysed (Schuster, 2008).

### *Mapping*

Linkage is the tendency of genes (or markers) to be inherited together because of their physical proximity to each other; the analysis of linkage between markers is the base for the construction of genetic maps and the subsequent molecular dissection of quantitative traits using positional information. A linkage map represents a “road map” of the chromosomes derived from two different parents (Paterson *et al.*, 1991), indicating the position and relative genetic distances between markers and trait loci along chromosomes. Dense genetic maps based on molecular markers provide a starting point for gene and QTL mapping, since they provide the framework in which the gene/QTL can be localized, and in some cases the interval of the chromosome in which marker/phenotype association is stronger. The first step in construction of a linkage map is the choice of parents that differ in one or many traits of interest. In order to maximize the number of heterozygous markers segregating in the population, genetically diverse parents need to be selected based for example on different pedigree origins.

This is particularly critical in the case of F1 populations, where genotypic evaluation using molecular markers will generate two maps (one for each parent) and a high proportion of heterozygous markers is required to achieve adequate genome coverage and provide statistical support for QTL analysis. Markers that are heterozygous in both parents can be used to anchor the two parental maps, which may be merged when such markers are in sufficient numbers. Markers for linkage map construction can also be selected on the basis of already known chromosomal positions. After selecting markers

based on heterozygosity in the parents and/or position on reference maps, the whole population is genotyped to obtain segregation data. These will finally be used to calculate recombination frequencies and construct the map.

One of the main limitations in peach mapping is its low level of genetic variability (Byrne *et al.*, 1991; Mnejja *et al.*, 2010), which results in a high proportion of the molecular markers assayed in a particular progeny being monomorphic.

To overcome this problem, some maps were produced using F2 progenies from interspecific almond × peach crosses that were highly polymorphic.

One of the most important is the *Prunus* Reference Map (Joobeur *et al.*, 1998; Aranzana *et al.*, 2003; Dominguez *et al.*, 2003; Dirlewanger *et al.*, 2004). It has been constructed using an F2 population of 111 seedlings from the MB1-73 Texas × Earlygold (almond × peach) hybrid (Joobeur *et al.*, 1998) and includes 536 markers; the wide genetic divergence between the parents means that this cross is highly polymorphic and the corresponding map provides excellent coverage of the 8 chromosomes of the *Prunus* genus, with a total distance of 519 cM and average density of 0.92 cM per marker.

In 2005 a “bin mapping” strategy that only uses a reduced set of individuals was developed. It consists of six selected plants (numbers 5, 12, 23, 30, 34 and 83) of the T × E F2 mapping population and two of their ancestors: Earlygold (E) and the F1 hybrid (H). This reduced bin mapping is representative of the recombination events in the progeny and is a powerful tool to quickly and efficiently map markers in T × E (Howad *et al.*, 2005).

All maps, including the most recently developed in different *Prunus* populations (Folta *et al.*, 2009), contain a framework of markers in common with T × E that allows identification of the linkage groups and ensures good coverage and marker spacing of the genome.

Over the last two decades, the availability of genetic knowledge on peach, considered a model for *Prunus* and the Rosaceae, has accelerated leading to the development of molecular markers, linkage and physical maps, comparative genomics studies. The last step has been the completion and public release of 8× whole-genome sequence on 1 April 2010.

Table 3. Peach major genes and QTLs affecting morphological or agronomic characters that have been on the *Prunus* reference map (Arús et al., 2012)

Characters	LGa	Symbol <sub>b</sub>	Populations	References
Flesh color (white/yellow)	G1	Y	'Padre' × '54P455'	Warburton et al. (1996); Bliss et al. (2002)
Evergrowing	G1	Evg	'Empress op op dwarf' × PI442380	Wang et al. (2002)
Internode length	G1	QTL	( <i>P. ferganensis</i> × 'IF310828')BC1	Verde et al. (2002)
Powdery mildew resistance	G1	QTL	'Summergrand' × P1908	Foulongne et al. (2003)
Flower color	G1	B	'Garfi' × 'Nemared'	Jauregui (1998)
PPV resistance	G1	QTLs	'Summergrand' × P1908; 'Summergrand' × P1908 F2; 'Rubira' × P1908	Decroocq et al. (2005); Marandel et al. (2009); Rubio et al. (2010)
Chilling and Heat requirement, Blooming date	G1	QTLs	'Contender' × 'Fla.92-2C'	Fan et al. (2010)
Root-knot nematode resistance	G2	Mi <sup>c</sup>	'P.2175' × 'GN22,' 'Akame' × 'Juseitou,' 'Lowell' × 'Nemared,' 'Garfi' × 'Nemared,' 'Padre' × '54P455'	Claverie et al. (2004); Yamamoto et al. (2001); Lu et al. (1998); Bliss et al. (2002); Jáuregui (1998)
Ripening time, fruit skin color, soluble-solids content	G2	QTLs	( <i>P. ferganensis</i> × 'IF310828')BC1	Verde et al. (2002)
Double flower	G2	Dl	'NC174RL' × 'PI'	Chaparro et al. (1994)
Broomy (or pillar) growth habit	G2	Br	Various progenies	Scorza et al. (2002)
PPV resistance	G2	QTLs	'Summergrand' × P1908; 'Rubira' × P1908	Decroocq et al. (2005); Rubio et al. (2010)
Blooming date	G2	QTL	'Contender' × 'Fla.92-2C,' 'Summergrand' × P1908	Fan et al. (2010); Quilot et al. (2004)
Flesh color around the stone	G3	Cs	'Akame' × 'Jusetou'	Yamamoto et al. (2001)
Anther color (yellow/anthocyanic)	G3	Ag	'Texas' × 'Earlygold'	Joobeur (1998)
Leaf curl resistance	G3	QTL	'Summergrand' × P1908	Viruel et al. (1998)
Fruit weight, fruit diameter, glucose content	G3	QTLs	'Suncrest' × 'Bailey'	Abbott et al. (1998)
Polycarpel	G3	Pcp	'Padre' × '54P455'	Bliss et al. (2002)
Flower color	G3	Fc	'Akame' × 'Jusetou'	Yamamoto et al. (2001)
Blooming time, ripening time, fruit development period	G4	QTLs	'Ferjalou Jalousia®' × 'Fantasia'; ( <i>P. ferganensis</i> × 'IF310828')BC1; 'Venus' × 'BigTop'; 'Summergrand' × P1908	Etienne et al. (2002); Verde et al. (2002); Cantin et al. (2010); Quilot et al. (2004)
Soluble-solid content, fructose, glucose	G4	QTLs	'Ferjalou Jalousia®' × 'Fantasia'; 'Venus' × 'BigTop'; 'Summergrand' × P1908	Etienne et al. (2002); Cantin et al. (2010); Quilot et al. (2004)
Flesh adhesion (clingstone/freestone)	G4	F	( <i>P. ferganensis</i> × 'IF310828')BC1; 'Akame' × 'Juseitou'	Verde et al. (2002); Dettori et al. (2001); Yamamoto et al. (2001)
PPV resistance	G4	QTLs	'Summergrand' × P1908; 'Summergrand' × P1908 F2; 'Rubira' × P1908	Decroocq et al. (2005); Marandel et al. (2009); Rubio et al. (2010)
Chilling requirement, blooming date	G4	QTL	'Contender' × 'Fla.92-2C'	Fan et al. (2010)
Chilling injury traits	G4	QTL	'Venus' × 'BigTop'	Cantin et al. (2010)
Fruit size	G4	QTLs	'Venus' × 'BigTop'; 'Summergrand' × P1908	Cantin et al. (2010); Quilot et al. 2004
pH, titratable acidity	G4	QTLs	'Venus' × 'BigTop'	Cantin et al. (2010)
Plant height	G4	QTL	'Venus' × 'BigTop'	Cantin et al. (2010)
Non-acid fruit	G5	D	'Ferjalou Jalousia®' × 'Fantasia'	Dirlewanger et al. (1998, 1999); Etienne et al. (2002)

Table 3 (continued)

Characters	LG <sub>a</sub>	Symbol <sub>b</sub>	Populations	References
Sucrose, malate, titrable acidity, pH, sucrose	G5	QTLs	'Ferjalou Jalousia@' × 'Fantasia'; 'Summergrand' × P1908 'Ferjalou Jalousia@' × 'Fantasia'; 'Padre' ×	Etienne et al. (2002); Quilot et al. (2004)
Skin hairiness (nectarine/peach)	G5	G	'54P455'	Dirlewanger et al. (1998, 1999); Bliss et al. (2002)
Kernel taste (bitter/sweet)	G5	Sk	'Padre' × '54P455'	Bliss et al. (2002)
PPV resistance	G5	QTLs	'Summergrand' × P1908 F2; 'Rubira' × P1908	Marandel et al. (2009); Rubio et al. (2010)
Chilling requirement, blooming date	G5	QTLs	'Contender' × 'Fla.92-2C'	Fan et al. (2010)
Ripening time, fruit skin color, soluble-solids content	G6	QTLs	(P. ferganensis × 'IF310828')BC1	Verde et al. (2002)
Plant height (normal/dwarf)	G6	Dw	'Akame' × 'Juseitou'	Yamamoto et al. (2001)
Leaf shape (narrow/wide)	G6	NI	'Akame' × 'Juseitou'	Yamamoto et al. (2001)
Male sterility	G6	Ps	'Ferjalou Jalousia@' × 'Fantasia'	Dirlewanger et al. (1998, 2006)
Powdery mildew resistance	G6	Vr2	'Rubira' × 'Pamirskij 5' F2	Pascal et al. (2010)
Powdery mildew resistance	G6	QTL	'Summergrand' × P1908	Foulongne et al. (2003)
Leaf curl resistance	G6	QTL	'Summergrand' × P1908	Viruel et al. (1998)
Fruit shape (flat/round)	G6	S*	'Ferjalou Jalousia@' × 'Fantasia'	Dirlewanger et al. (1998, 1999, 2006)
Aborting fruit	G6	Af	'Ferjalou Jalousia@' × 'Fantasia'	Dirlewanger et al. (2006)
PPV resistance	G6	QTLs	'Summergrand' × P1908; 'Summergrand' × P1908 F2; 'Rubira' × P1908	(2010) Decroocq et al. (2005); Marandel et al. (2009); Rubio et al.
Chilling requirement, blooming date	G6 G6-	QTLs	'Contender' × 'Fla.92-2C'	Fan et al. (2010)
Leaf color (red/yellow)	G8 G6-	Gr	'Garfi' × 'Nemared'; 'Akame' × 'Juseitou'	Jauregui (1998); Yamamoto et al. (2001)
Fruit skin color	G8	Sc	'Akame' × 'Juseitou'	Yamamoto et al. (2001)
Leaf gland (reniform/globose/eglandular)	G7	E	(P. ferganensis × 'IF310828')BC1	Dettori et al. (2001)
Resistance to mildew	G7	QTL	(P. ferganensis × 'IF310828')BC1	Verde et al. (2002)
PPV resistance	G7	QTLs	'Summergrand' × P1908; 'Summergrand' × P1908 F2; 'Rubira' × P1908	(2010) Decroocq et al. (2005); Marandel et al. (2009); Rubio et al.
Chilling requirement, blooming date	G7	QTLs	'Contender' × 'Fla.92-2C'	Fan et al. (2010)
Flower morphology	G8	Sh	'Contender' × 'Fla.92-2C'	Fan et al. (2010)
Powdery mildew resistance	G8	QTL	'Summergrand' × P1908	Foulongne et al. (2003)
Quinase	G8	QTL	'Ferjalou Jalousia@' × 'Fantasia'	Etienne et al. (2002)
Chilling requirement, heat requirement	G8	QTLs	'Contender' × 'Fla.92-2C'	Fan et al. (2010)

### *The peach genome*

*Prunus persica*, unlike other temperate fruit crops, such as European plum, sour cherries, apple and pear, that are of polyploid origin, is a diploid with  $n=8$  and a relatively small genome equivalent to ~300 Mbp per haploid genome (Baird et al. 1994), about twice that of *Arabidopsis thaliana* (Arumuganathan & Earle, 1991). The use of heterozygous varieties could be problematic for genomic sequence assembly. To overcome this issue a doubled haploid derived from the cultivar Lovell (Toyama, 1974), was employed for genome sequencing. Lack of allelic variation in this homozygous genotype allowed a deeper effective genome coverage and greatly simplified the assembly of the genome (Arús et al., 2012).

To process the scaffolds for appropriate chromosomal assignment, anchoring, and orienting, an updated version of the Texas × Earlygold *Prunus* reference map was used. The fact that peach genome was generated entirely from pair end reads of Sanger sequencing, together with these features, allowed to generate a high-quality chromosome scale assembly with a very high level of anchoring and orienting, to chromosomes (Arús et al., 2012).

The paper describing the high quality draft genome sequence of peach is going to be published by the International Peach Genome Initiative (Verde et al in press.).

The availability of the peach genome sequence and the efficiency of next generation sequencers has made affordable the whole genome resequencing, facilitating the genome-scale SNP discovery.

In 2011 the International Peach SNP Consortium (IPSC) resequenced 56 peach accessions identify genome-wide sequence variation and to develop an high-throughput SNP genotyping platform based on the Illumina Infinium® technology (Verde et al., 2012 Figure 10, Figure 11).

The IPSC peach SNP array includes 8,144 SNP (out of an identified total of 1,022,354) distributed all over the genome. When tested on a pool of 709 accessions, over the 80% of them resulted to be polymorphic (Verde et al., 2012).

The IPSC peach 9K SNP array v1 is commercially available and represent an high-throughput, high-resolution and cheap method to genotype populations and to obtain molecular markers linked to genes involved in the expression of important agronomic traits (both monogenic and polygenic) that can be used for marker-assisted selection (MAS).

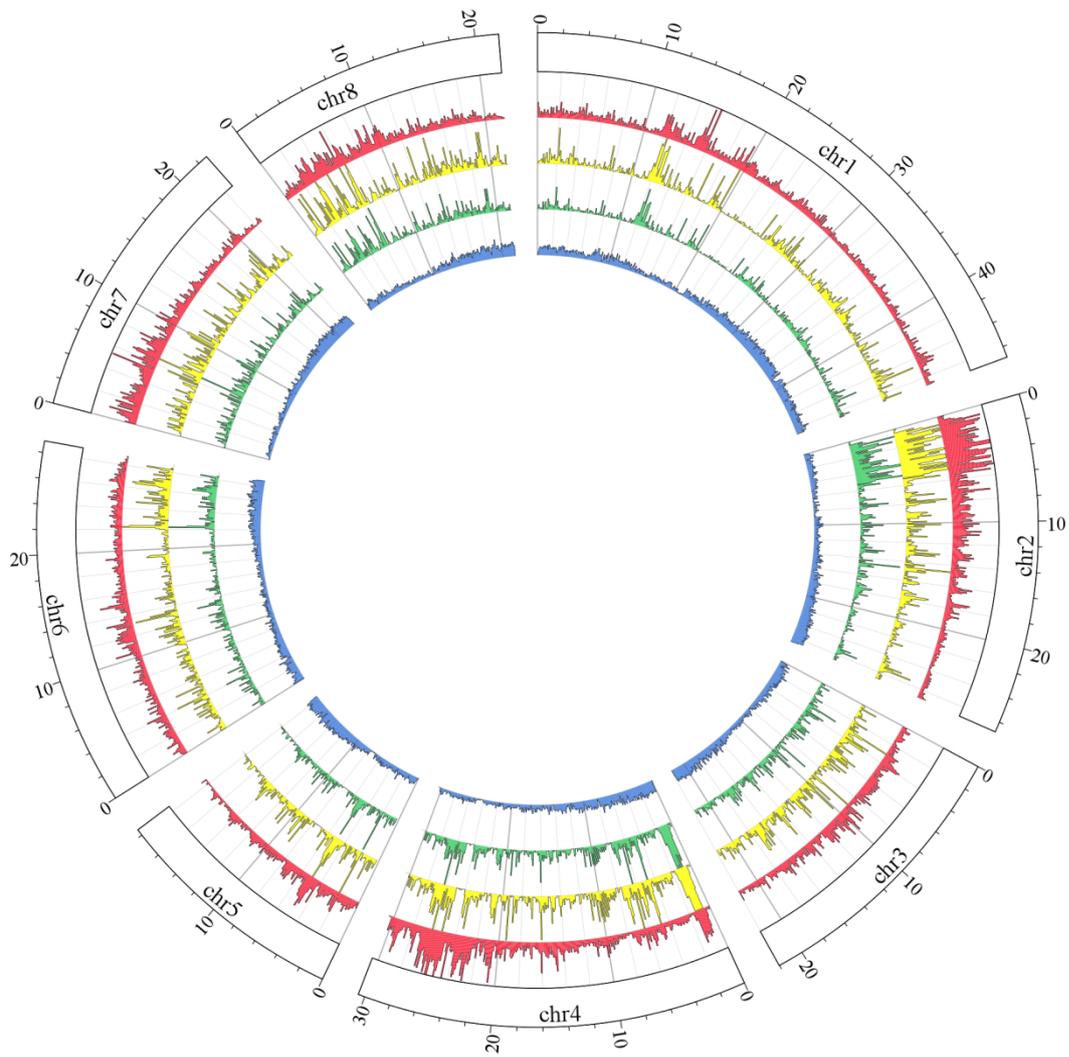


Figure 10. Distribution of SNPs along the Peach v1.0 pseudomolecules (Verde et al., 2012)

Pool	Accession	Adaptors	Read length (bp)	Read count (million)	Coverage of peach genome
1	'Armkung'	CACAGT	94	5.85	2.42
1	'Big Top'	CGAGAT	94	3.55	1.47
1	'Fidelia'	ATGGCT	94	6.47	2.68
1	'Flordastar'	GCATAG	94	7.27	3.01
1	'Silver Rome'	CATTCG	94	8.35	3.45
1	'Weinberger'	ACACTG	94	9.72	4.02
2	'Babygold 8'	TTGCGA	93	5.60	2.29
2	'Elberta'	CAGTAC	93	5.63	2.30
2	'Maruja'	TGCAAC	93	8.52	3.49
2	'Maycrest'	ACTAGC	93	8.61	3.52
2	'Oro A'	GAGCAA	93	7.20	2.95
2	'Stark Red Gold'	GCTACA	93	6.37	2.61
3	'Circe'	CATTCG	93	9.23	3.78
3	'Imera'	GCATAG	93	5.92	2.42
3	'Percoca di Romagna 7'	ATGGCT	93	4.27	1.75
3	'Pillar'	ACACTG	93	1.40	0.57
3	'S 2678'	CGAGAT	93	10.15	4.15
3	'Stark Saturn'	CACAGT	93	7.45	3.05
4	'Kamarat'	ACTAGC	93	9.63	3.94
4	'Leonforte 1'	GAGCAA	93	2.32	0.95
4	'Sahua Hong Pantao'	GCTACA	93	19.20	7.86
4	'Shen Zhou Mitao'	TTGCGA	93	12.54	5.13
4	'Tabacchiera'	TGCAAC	93	0.56	0.23
4	'Tudia'	CAGTAC	93	7.43	3.04
5	'GF677' <sup>1</sup>	CGAGAT	93	9.22	3.77
5	'Kurakata Wase'	ATGGCT	93	6.75	2.76
5	'Quetta'	CACAGT	93	12.76	5.22
5	'S6699'	ACACTG	93	4.90	2.01
6	'Admiral Dewey'	GGGT	80	2.42	0.85
6	'Babcock'	CCAT	80	3.19	1.12
6	'Elberta'	AGCT	80	0.64	0.23
6	'Slappey'	TCCT	80	2.02	0.71
7	'Bolinha'	AGCT	80	3.55	1.25
7	'Carmen'	GGGT	80	1.66	0.58
7	'Chinese Cling'	TCCT	80	2.50	0.88
7	'Mayflower'	CCAT	80	1.35	0.47
8	'Diamante'	AGCT	80	2.11	0.74
8	'J.H. Hale'	TCCT	80	3.18	1.12
8	'Rio Oso Gem'	CCAT	80	2.57	0.91
8	'Yellow St. John'	GGGT	80	1.35	0.48
9	'Dixon'	GGGT	80	1.25	0.44
9	'Early Crawford'	TCCT	80	3.89	1.37
9	'Florida Prince'	CCAT	80	1.85	0.65
9	'Nonpareil' <sup>2</sup>	AGCT	80	2.52	0.89
10	'Dr. Davis'	GGGT	80	2.31	0.81
10	'Nemaguard'	AGCT	80	2.38	0.84
10	'O'Henry'	TCCT	80	4.28	1.51
10	'Okinawa'	CCAT	80	2.15	0.76
11	'Georgia Belle'	AACT	80	14.42	5.08
11	'Lovell'	GTGT	80	6.55	2.30
11	'Lovell'	CCTT	80	0.03	0.01
11	'Oldmixon Free'	TGGT	80	3.26	1.15
12	'Big Top'	ACACGTAGTAT	330	0.20	0.29
12	'Binaced'	ACACGAGACT	355	0.16	0.26
12	'Catherina'	ACACTACTCGT	288	0.17	0.22
12	'Elegant Lady'	ACGACACGTAT	243	0.19	0.20
12	'Nectaross'	ACGAGTAGACT	275	0.19	0.23
12	'O'Henry'	ACGCGTCTAGT	289	0.15	0.18
12	'Sweet Cap'	ACGTACACACT	251	0.16	0.18
12	'Venus'	ACGTACTGTGT	278	0.15	0.19

<sup>1</sup>Peach x almond hybrid;

<sup>2</sup>Almond accession.

Pools 1–11 were sequenced with the Illumina Genome Analyzer while pool 12 was sequenced with the Roche 454 platform. Adaptors were used for retrieving accession-specific sequences from pools.

doi:10.1371/journal.pone.0035668.t001

Figure 11. Accessions of peach, almond and peach x almond hybrid sequenced at the Istituto di Genomica Applicata (IGA, Udine, Italy) (pools 1–5), the Center for Genome Research and Biocomputing (CGRB, Oregon State University, Corvallis, OR, USA) (pools 6–11), and IRTA (Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB, Spain) (pool 12).

The genomic localization of many traits have been determined on the peach linkage maps. Moreover, some genes controlling important traits have been fine-mapped and their map-based cloning is underway.

NGS-based methods are already starting to enhance classical candidate-gene approaches such as genetic mapping and positional cloning by making them faster and more efficient and enabling the investigation of variations, including SNPs, small indels, and structural variants in the germplasm to identify genes and alleles of interest.

In *Arabidopsis thaliana* the next-generation sequencing (with a 22x coverage) of a slow-growing and reduced-pigmentation induced mutant, was used to identify a small interval associated with the mutation that, when manually scanned, contained the causal mutation (Schneeberger *et al.*, 2009).

The availability of the peach genome sequence has allowed moreover the comparison with the full sequence of other rosaceous genomes. In a recent work the apple genome (Velasco *et al.* 2010) was compared with those of diploid strawberry (Shulaev *et al.*, 2011) and peach to derive the ancestral genome structure and to study the genome evolution in these species. The analysis of 806 markers (of which 784 in common with peach) allowed to hypothesize some of the features of an ancestral nine chromosomes genome, to reveal patterns of conservation of synteny across the genomes of the three genera and to discover trends in the evolution of a group of closely related species with important differences in plant habit, reproductive behavior, and fruit characteristics. Moreover the comparison of peach and apple sequencing data allowed to define the genomic reallocation of chromosomal parts that occurred during speciation. This approach made it also possible to hypothesize the genome of the ancestral species from which both apple and peach originated. This study simplified the process of identifying the syntenic regions between apple and peach and transferring molecular information between them.

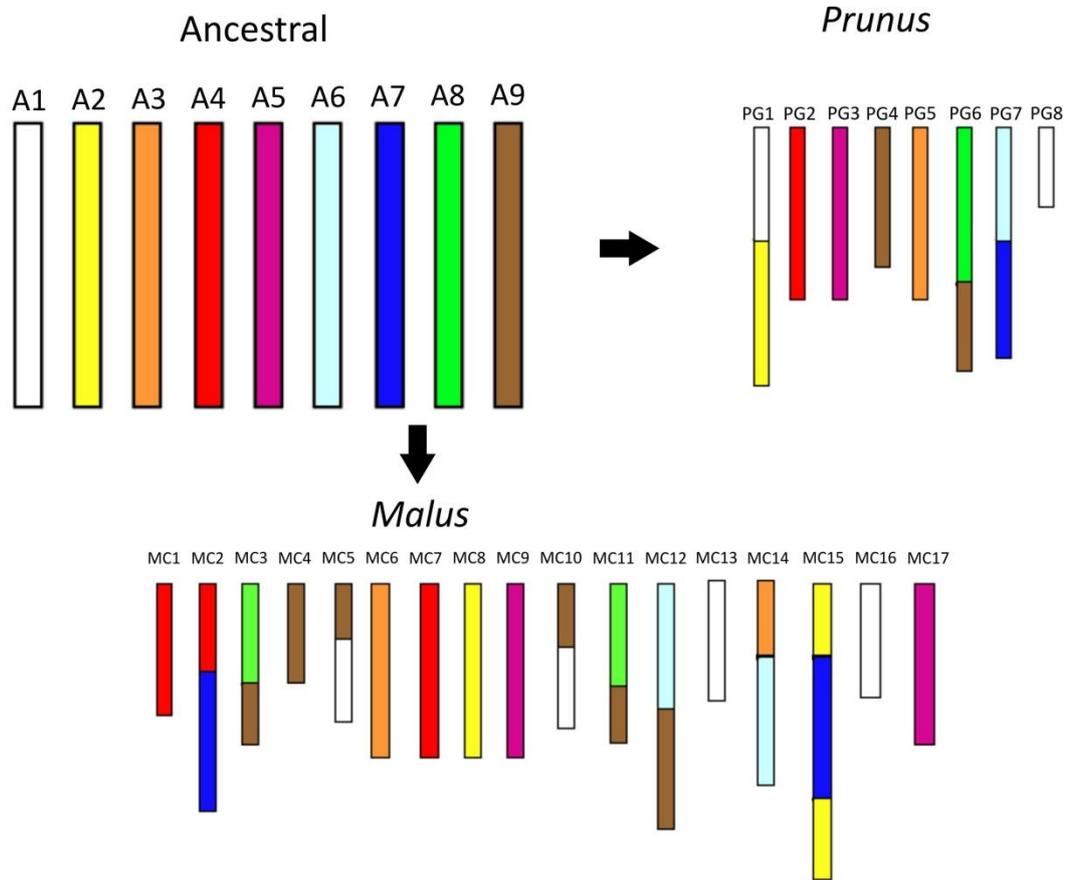


Figure 12 Syntenic regions among the genomes were elucidated from the positions of 129 orthologous markers shared by all three genomes. The hypothetical ancestral genome contains nine chromosomes numbered Ancestral 1 (A1) - A9. Sections of the chromosomes of and Prunus are coloured according to the hypothetical ancestral chromosomes. Extant chromosome/linkage group lengths assume that all nine hypothetical ancestral chromosomes were of the same length. (Modified from Illa et al., 2011)



## Aim of the work

In plants, the differences in carotenoid accumulation can be due to three different mechanisms: biosynthesis, compartmentalization and degradation. In the case of peach, the yellow flesh color is determined by the strong accumulation of carotenoids in fruit cells. On the other hand, white peaches are reported to contain very limited amounts of carotenoids but conversely, they are able to produce large amounts of apocarotenoids that are derived from carotenoid degradation. A strong expression of a carotenoid dioxygenase gene (*ccd4*) is reported in white in respect to the yellow peach flesh. Finally, the Y locus controlling the peach flesh color was preliminarily mapped on the LG1 of peach. Starting from these knowledge, we undertake a research project aimed at identifying the gene controlling the yellow/white fruit flesh color in peach.

To reach this goal the genomic region from scaffold 1 included between the closest markers flanking the Y locus (Brandi, 2010) will be analyzed *in silico* to predict open reading frames, among which functional candidate genes for the control of carotenoid metabolism in fruit flesh will be identified.

The best candidates found within this region will be fully sequenced in different genotypes in order to identify possible allelic variants linked to the white and yellow phenotypes. Within the pool of genotypes two spontaneous mutants, White Redhaven (sport mutant of the yellow variety Redhaven) and Cristina (yellow mutant of the white cultivar Caldesi 2000) will be deeply analyzed including chimerical tissues, which are easily recognizable in correspondence of the fruit suture.

The association between allelic variants and flesh color will be also verified in a pool of varieties from the modern and traditional peach germplasm and in a cross progeny segregating for flesh color. Different approaches for the genetic and functional validations will be evaluated. In detail, sequencing information will be used to develop highly specific markers to be used for the mapping in segregating progenies and for the early marker assisted selection in future breeding programs. The allelic variability of the candidate gene will be discussed in relation to the evolution of peach flesh color. Then a construct for the heterologous expression of the candidate gene will be prepared to open the way of a deeper functional analysis.



## **Material and Methods**



## Plant material

In this work several accessions of peach germplasm have been used. Most of them were harvested from trees grown in the experimental farm of the Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Unità di Ricerca per la Frutticoltura di Forlì (CRA FRF) in the Po Valley.

Two pairs of mutant sports were analyzed in this work: Redhaven and its mutant sport White Redhaven; Caldesi 2000 and its sport mutation Cristina. The pedigrees of the original cultivars are reported in Figure 13.

Embryos derived from self-fertilization of Redhaven and White Redhaven were also used for candidate genes sequencing.

The 84 individuals used for mapping derive from the cross Royal Prince × Yoshihime. The population segregates for flesh color, being the fruits of the two parents yellow and white fleshed respectively.

The 96 white and yellow-fleshed genotypes, representative of the peach germplasm variability include worldwide released varieties, CRA FRF Advanced Selections and, old varieties of the heritage Italian germoplasm (Table 4).

The dried leaves of the peach varieties Admiral Dewey, Early Crawford, Elberta, J.H.Hale, Kalamazoo, Muir, Rio Oso Gem, Yellow St. John, Chinese Cling, Georgia Belle, (founder varieties of American breeding programs) were kindly supplied by the National Clonal Germplasm Repository for Fruit and Nut Crops, Davis (CA). In total, 106 peach accessions have been analyzed (Table 4).

**Worldwide released varieties**

Adriana	Y	Alba	W
Alired	Y	Aliblanca	W
Alitop	Y	Alipersiè	W
Alix	Y	Amanda	W
Ambersisters	Y	Artic Sweet	W
Andross	Y	Benedicte	W
Azurite	Y	Caldesi 2000	W
Babygold 9	Y	Crizia	W
Big Top	Y	Douceur	W
Copia Poa	Y	Early Giant	W
Coraline	Y	Early Silver	W
Cristina <sup>b</sup>	Y	Emeraude	W
Diamond Ray	Y	Gladis	W
Egea	Y	Greta	W
Eolia	Y	Jade	W
Guglielmina	Y	Kurakatawase	W
Honey Blaze	Y	Maria Anna	W
Honey Kist	Y	Maria Delizia	W
Jonia	Y	Maylis	W
Jungerman	Y	Neve	W
Kawah	Y	Pearlsisters D93 1-19	W
Lady Erica	Y	Platfortwo	W
Maria Aurelia	Y	Sahong	W
Maria Dolce	Y	September Snow	W
Maria Dorata	Y	Silver Giant	W
Max 7	Y	Silver Rome	W
Maycrest	Y	Snow Queen	W
Nectaross	Y	Spring Snow	W
O'Henry	Y	Spring White <sup>c</sup>	W
Oro A	Y	Stark Saturn	W
Red Top	Y	Summer Sweet	W
Red Valley	Y	Tendresse	W
Redhaven	Y	Vanilia	W
Rich Lady	Y	White Redhaven <sup>d</sup>	W
Rome Star	Y	Yoshihime	W
Rose Diamond	Y	Yumyeung	W
Royal Glory	Y	Zephir	W
Royal Prince	Y		
Royal Summer	Y		
Rubirich	Y		
September Free	Y		
Valley Red	Y		
Velvetsisters	Y		
Vistarich	Y		
Zee Lady	Y		

**Heritage Italian Germoplasm**

Fuoco di Romagna	Y
Gialla Tardiva	Y
Percoca di Romagna	Y
Percoca di Romagna 7	Y
Pesca Carota	Y
Bella di Cesena	W
Bella di Cesena precoce	W
Bella di Piangipane	W
Buco Incavato	W
Iris Rosso	W
Rosa del West	W
S.Anna Balducci	W

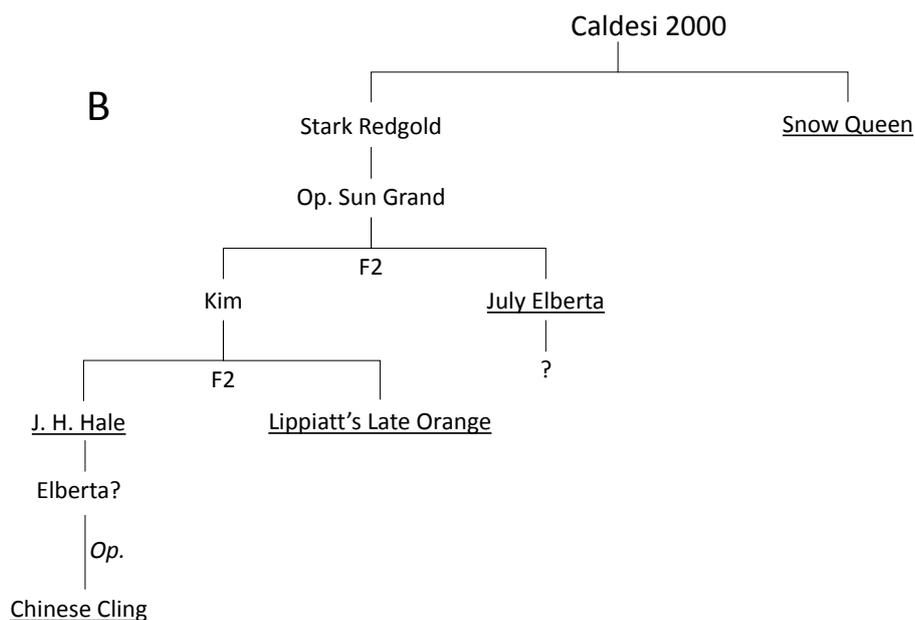
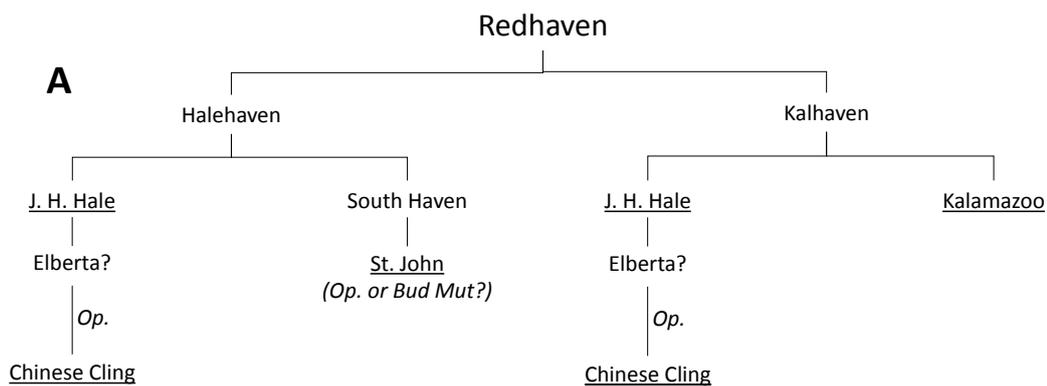
**US breeding founders**

Admiral Dewey	Y
Early Crawford	Y
Elberta	Y
J.H.Hale	Y
Kalamazoo	Y
Muir	Y
Rio Oso Gem	Y
Yellow St. John	Y
Chinese Cling	W
Georgia Belle	W

**CRA FRF Advanced Selections**

IFF 331	W
IFF 813	Y

Table 4. List of peach accessions analyzed in this work.



**C**

Variety	Type	Origin	Year
Caldesi 2000	nwf	Italy	1984
Stark Redgold	nyf	California	1962
Snow Queen	nws	California	1975
Sun Grand	nyf	California	1950
Kim	nyf	California	1938
July Elberta	pyf	California	1930
J. H. Hale	pyf	Connecticut	1912
Lippiatt's Late Orange	nyf	New Zealand	1916
Elberta	pyf	Georgia	1889
Redhaven	pys	Michigan	1940
Halehaven	pyf	Michigan	1932
Kalhaven	pyf	Michigan	1936
South Haven	pyf	Michigan	1916
St. John	pyf	USA	1860
Kalamazoo	pyf	Michigan	1869

Figure 13. Pedigree of the Redhaven (A) and Caldesi 2000 (B) cultivars and ancestors' origin (C)

## DNA extraction

### *DNA extraction from leaf*

Young leaves harvested from actively growing shoots were used for DNA extraction. The central rib was removed and the leaf blade inserted in a 2 ml PP eppendorf tube. After freezing at  $-80^{\circ}\text{C}$ , leaves within the tubes were lyophilized for 36hr. A pinch of silicon carbide was then added to each tube and leaves were finely grinded for 3 minutes using Mixer mill (Retsch GmbH, Haan, Germany) set at 29hz. Powdered leaves were then stored at  $-20^{\circ}\text{C}$  until use. The protocol is as follows;

- Add 900  $\mu\text{l}$  of CTAB buffer (see Table 5) preheated at  $65^{\circ}\text{C}$  to 5mg of leaf powder in a 2ml eppendorf tube.
- Vortex the sample thoroughly and incubate for 30 minutes  $65^{\circ}\text{C}$
- Add 900  $\mu\text{l}$  of dichloromethane-isoamyl alcohol 24:1 and thoroughly mixed until complete emulsion
- Centrifuge the sample at 5000 rpm for 5 minutes
- Transfer the supernatant in a new tube and add an equal volume of cold ( $-20^{\circ}\text{C}$ ) isopropanol. To ensure a better precipitation of nucleic acid, the sample can be left for some time (i.e. 20 minutes to overnight) at  $-20^{\circ}\text{C}$ .
- Centrifuge the sample at 13000 rpm for 5 minute, remove supernatant and clean the pellet using 500  $\mu\text{l}$  of washing buffer (see Table 5)
- Dry the pellet and resuspend in 200  $\mu\text{l}$  of sterile water

<b>CTAB Buffer</b>	
CTAB (Sigma)	2 %
NaCl	1,4 M
Tris-HCl pH8	100 mM
EDTA	20 mM
PVP-40	2 %
$\beta$ -mercapto-ethanol	1 %

<b>Washing Buffer</b>	
Ethanol	76 %
Ammonium acetate	10 mM

*Table 5. DNA extraction solutions*

#### *DNA extraction from embryos*

Peach seed was opened and the embryo carefully cleaned from the maternal coating and cut in half for the extraction. The remaining half was put at -20°C for long term storage.

- Grind the half of the seed in liquid nitrogen and transfer the powder inside a 2 ml eppendorf tube
- Add 900  $\mu$ l of CTAB buffer (see Table 5) preheated at 65°C to the tube containing the material
- Vortex the sample thoroughly and incubate for 30 minutes at 65°C
- Add 900  $\mu$ l of phenol-chloroform-isoamyl alcohol 25:24:1 and thoroughly mixed until complete emulsion
- Centrifuge the sample at 5000 rpm for 5 minutes
- Transfer the supernatant in a new tube, re-add 900  $\mu$ l of phenol-chloroform-isoamyl alcohol 25:24:1 and thoroughly mix until complete emulsion
- Centrifuge the sample at 5000 rpm for 5 minutes
- Transfer the supernatant in a new tube and add an equal volume of cold (-20°C) isopropanol. To ensure a better precipitation of nucleic acid, the sample can be left for some time (i.e. 20 minutes to overnight) at -20°C.
- Centrifuge the sample at 13000 rpm for 5 minute, remove surnatant and clean the pellet using 500  $\mu$ l of washing buffer (see Table 5)

- Dry the pellet and resuspend in 200 µl of sterile water

Quality and concentration of extracted samples were assayed using a Nanodrop ND-1000 spectrophotometer (Nanodrop technologies, Wilmington, Delaware USA).

## PCR

### *Standard amplifications*

Standard PCR amplifications of microsatellites and fragments sized up to 2kb were conducted in an automated thermal cycler (MJ Research DNA Thermal cycler PTC-200) with AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Foster City, CA, USA); PCR was performed in 20µl mixtures containing 100µg genomic DNA, 2.5mM MgCl<sub>2</sub>, 0.3 mM each dNTP, 0.5 µM each primer and 2U Taq polymerase. The following conditions were used: denaturation at 94°C for 7 minutes; 35 cycles of: annealing at primer-specific temperature (see Table 6) for 1 minute, extension at 72°C for 2 minutes 30 seconds, denaturation at 94°C for 20 seconds; and final extension at 72°C for 10 minutes.

### *Insertion detection (Three primer system)*

To quickly evaluate the presence of the insertion, a three-primer amplification was set up using three primer: ccd4-EX1for, ccd4-E2rev, flanking the intron, and ccd4-INS1for, located at 3' end of the insertion. The amplification of the ccd4-EX1for forward primers with the shared reverse produce a fragments of 352 bp when an allele without insertion is present and gives no product in the opposite case because of the large size of insertion in between. Ccd4-INS1for oligo gives a 263 bp product only when the retrotransposon is present.

Thus, this system allow a fast and reliable method to assay the allele composition of *ccd4*. PCR is performed in 15µl mixtures containing 50µg genomic DNA, 2.5mM MgCl<sub>2</sub>, 0.3 mM each dNTP, 0,3 µM ccd4-EX1for and ccd4-E2rev primer each, 0,5 µM ccd4-INS1for and 1,3U Taq polymerase. The following conditions were used: denaturation at 94°C for 7 minutes; 35 cycles of: annealing at primer-specific temperature (see Table 6)

for 45 seconds, extension at 72°C for 45 seconds, denaturation at 94°C for 20 seconds; and final extension at 72°C for 5 minutes.

#### *Amplification of long templates and Hi-Fidelity PCR*

Amplification of long templates and fragments for cloning were carried out using the Herculase® II Fusion DNA Polymerase (Agilent, Waldbronn, Germany). This enzyme is a high fidelity Pfu-based DNA polymerase that enables the amplification of templates up to 20kb long, using a modified amplification protocol. Amplifications were carried out in 50µl mixtures containing 150 ng DNA template, 2.5 mM MgCl<sub>2</sub>, 3% DMSO, 0.3 mM each dNTP, 0.2 µM each primer and 2.5U DNA polymerase. The cycling conditions were 95°C for 4 minutes, followed by 35 cycles of: 92°C for 20 seconds, 60°C for 40 seconds, 69°C for 7 minutes, and finally 72 °C for 8 minutes. The protocol adopted for PCR-cloning is the same, with the exception of elongation time and temperature: 30 sec/kb at 72°C.

#### *Temperature switch PCR (TSP)*

TSP is a special type of PCR designed for screening of single nucleotide polymorphisms (SNPs). Each reaction contains three primers with fundamentally different melting temperatures (TM): the locus specific (TSP-LocF/TSP-LocR; Table 6) primers with a relatively high TM, and the allele specific (TSP-AF; Table 6) primer with a relatively low TM (Figure 14) designed with a short, non-complementary 5' tail aimed to raise the melting temperature.

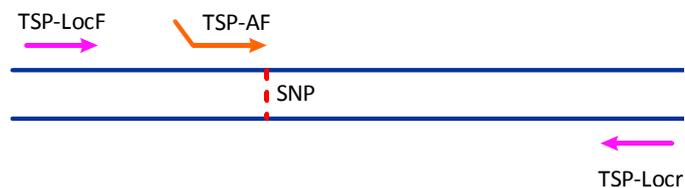


Figure 14. Primer design for TSP (adapted from Hayden et al., 2009)

In the first part of the PCR the outer TSP-Loc primers are used to amplify the ~500bp target locus from genomic DNA. Then, a drop in the PCR annealing temperature (hence the name TSP) switches the focus to the inner locus specific primer. The TSP-AF primer contains the mutated base at its 3' terminus. This primer only binds to the mutant allele, and thus the 'allele specific' product is only amplified if the mutation is present.

TSP PCR was performed in 16µl mixtures containing 50 ng DNA template, 1.5 mM MgCl<sub>2</sub> 0.2 mM each dNTP, 0.1 µM each locus specific primer, 0.5 µM the allele specific primer, 100 ng/µl of BSA and 0.2U of Taq polymerase. Amplification was carried out for a total of 35 cycles with the following thermal profile: 30 seconds at 94°C, 30 seconds at 58°C, 60 seconds at 72°C for 15 cycles. The next five cycles were with 10 seconds at 94°C, 30 seconds at 45°C, followed by 15 cycles with 30 seconds at 94°C, 30 seconds at 53°C, 5 seconds at 72°C.

### *Primer design*

All primers were designed using Primer3 (<http://frodo.wi.mit.edu/primer3/input.htm>) on sequences retrieved from the peach genome or sequenced in the present study. During design process Primer3 was set to use a standard melting temperature between 58°C and 60°C to assure compatibility between different oligos.

Oligo	Sequence	Annealing	Lenght
ccd4FLfor	CCAAAACAGGGCAAACCTAA	58	20
ccd4-5for	GCAGTGAAGGGCAATACCAG	58	20
ccd4-SSRrev	TGGGCTTGATGGTCTTTCTT	58	20
ccd4-RACErev	AGTTTTTGGAGGCTGTGGTG	58	20
ccd4-EX1for	TCTTCCCAGCGTCTTCTCTG	58	20
ccd4-INS1rev	TTCACCCGTGACAGACAAAC	58	20
ccd4-E2rev	TGCCAAAGAAAGCCAACTT	58	20
ccd4-EX2rev	ATGGGTTGCCTTGACACAAT	58	20
ccd4-CDSrev	TATGTTGCTGAATGGGGACA	58	20
ccd4-FLrev	GCACTCACCTAGTTTGGGGTA	58	21
TSP-AllSpFor	CGCCAATGTGGAGCACT	49	17
TSP-LocSpFor	AATGCTTGGGATGAAGAGGA	58	20
TSP-LocSpRev	GAGGAGACTTGGCATCCATC	58	20

*Table 6. Oligo used for sequencing and genotyping of ccd4 alleles*

Oligo	Sequence	Annealing	Lenght
NDE5for	GTTTCATATGGAAGAAAGACCATCAAGCCCA	60	30
Eco3'rev	AACGAATTCTATGTTGCTGAATGGGGACA	61	29
pET28for	GCGGATAACAATTCCCCTST	58	20
pET28rev	AGTGGTGGTGGTGGTGGT	58	18

*Table 7. Oligo used for expression vector construction*

Oligo	Sequence	Annealing	Lenght
INS1rev	TTCACCCGTGACAGACAAAC	58	20
INS2rev	GGTTTTCTCCAAGCAAGTCTTT	58	22
INS3for	AATCCTTAAAGTGCCCAACG	58	20
INS4for	GCGCATGCATCAATTATAAGC	58	21
INS6for	GGTGGGAGCAGACAATAAGC	58	20
INS7rev	TTGAAATTCACCGTTCTACC	58	22
INS5rev	TCCTTTAAGTTTCTCAACCTCTG	58	24
INS8for	TGAATTGAGTAAAAGCCACGA	58	21
INS9for	TCTGGACTTCAAGCCCTTACA	58	21
INS10rev	CTCCGAGGACGTAGGCATAG	58	20
Eco0 ad.	GACTGCGTACCAATTC	46	16
M13 for	TGTA AACGACGGCCAGT	54	18
M13 rev	CAGGAAACAGCTATGACC	54	18

*Table 8 Oligo used for primer-walking of the insertion within White Redhaven.*

### *Agarose gel electrophoresis*

Agarose gel electrophoresis was used to separate DNA fragments bigger than 300bp. Samples were prepared by adding 0.2 volumes of Blue Dye (Table 9). Gel was prepared by melting agarose powder in 1x sodium borate (SB) buffer (Table 9). A standard agarose concentration of 1% w/v was usually adopted.

The samples were run for 45 minutes at 180 V. This unusually high voltage would normally cause overheating of gel, resulting in melting; however SB buffer is not prone to this phenomenon, allowing to use higher voltage during electrophoresis. Moreover SB buffer allows better resolution of low-weight band than the standard TAE buffer.

The gel was stained for 40 minutes in an ethidium bromide (EtBr) staining solution (Table 9) and images were acquired with a Kodak Image Station CF440 (Eastman Kodak Company, USA).

<b>Blue Dye</b>	
TAE 50x	2 ml
Ficoll 400	2,5 g
Bromophenol blue	25 mg
Distilled water	to 20 ml
<b>SB Buffer 20x</b>	
Boric acid	39,2 g
NaOH	8 g
Distilled water	up to 1 l
Adjust pH to 8,5 using NaOH pellets	
<b>EtBr staining solution</b>	
Ethidium Bromide 10mg/ml	100 µl
Distilled water	1l

*Table 9. Agarose electrophoresis solutions*

#### *Polyacrylamide gel electrophoresis*

Polyacrylamide gel electrophoresis was used to separate DNA fragments shorter than 400bp. Gel is prepared using TBE buffer added with urea (with a concentration of 7M to achieve denaturing conditions) bisacrylamide and acrylamide (Table 10). Polymerization occurs when polymerizing agent (APS, Ammonium persulfate) and catalyzer (N,N,N',N'-Tetramethyl-ethylenediamine, TEMED). Gel is formed by pouring the solution (Table 10) between two conveniently treated glass plates. Samples were prepared by adding 0.5 volumes of denaturing dye and heated to 95°C for 5'. The run is carried out at 65w constant power for 50 minutes to 2 hour, depending on the fragment size.

<b>TBE buffer 5x</b>	
Tris (Sigma)	54 g
Boric acid	27,5 g
EDTA 0,5 M pH 8	20 ml
Distilled water	up to 1l
<b>TBE – Urea</b>	
Urea (Sigma)	210 g
5x TBE	100 ml
Distilled water	up to 500 ml
<b>Polyacrylamide gel</b>	
TBE - Urea	55 ml
Liquacryl (40% acrylamide, MP Biomedicals)	8,5 ml
Liquabis (2% bis-acrylamide, MP Biomedicals)	8,5 ml
TEMED	45 µl
APS 10%	450 ul
<b>Denaturing loading buffer</b>	
EDTA 0,5 M pH 8	0,5 ml
Bromophenol blue	25 mg
Xylene cyanol	25 mg
Formamide 98%	up to 25 ml

*Table 10. Solutions for polyacrylamide gel electrophoresis*

Staining is made by silver nitrate. One glass plate is removed and the gel is soaked in Fix solution (Table 11) for 30 minutes. After two 5 minutes wash in distilled water, the gel is transferred to the staining solution (Table 11) for 45 minutes. Developing is preceded by a 20 second wash in distilled water and consists in two distinct soaking in developing solution (Table 11). When DNA bands reach the optimal contrast, the process is halted using the remaining fix solution.

<b>Fix solution</b>	
Acetic acid	200 ml
Distilled water	1,8 l
<b>Staining solution</b>	
AgNO <sub>3</sub>	1.5 g
Formaldehyde 37%	2 ml
Bidistilled water	up to 1 l
<b>Developing solution</b>	
Na <sub>2</sub> CO <sub>3</sub>	60 g
Formaldehyde 37%	3 ml
Sodium Thiosulfate 1%	400 µl
Bidistilled water	up to 2 l

Table 11. Polyacrylamide gel silver nitrate staining solutions

## Cloning

### *Ligase*

The pGEM t-easy vector system (Promega, Madison, Wisconsin USA) was used for cloning PCR products. The reaction was conducted on a final volume of 5 µl and incubated overnight at 4°C following manufacturer instructions.

### *A-tailing*

Taq DNA polymerases add a single adenine nucleotide to the 3' of the amplified fragment. This peculiarity is used by the t- cloning system: plasmids suitable for this technique are treated by adding a T to both the 5' ends of the linearized molecule. Thus the complementarity allows the direct cloning of PCR products without the need of a previous digestion to create compatible ends.

High fidelity enzymes, on the contrary, are characterized by a proof-reading activity and hence produce a blunt-ended fragment. To allow the T-cloning of such amplified fragments, the A-tailing procedure is necessary. It is carried on a blunt DNA fragment by a standard polymerase in presence of dATP. The process is carried out by incubating the A-tailing mix (see Table 12) at 72°C for 20 minutes.

<b>A-tailing</b>	
Purified PCR product	10 $\mu$ l
10X buffer	1,5 $\mu$ l
MgCl <sub>2</sub> 25 mM	0,6 $\mu$ l
dATP 10 mM	0,75 $\mu$ l
<i>Taq</i> -polymerase 5U/ $\mu$ l	0,2 $\mu$ l
Purified water	1,95 $\mu$ l
Total volume	15 $\mu$ l

*Table 12. A-tailing reaction mix*

### *Competent cells*

Two different *Escherichia coli* strains were used for chemical competent cells preparation: DH5 $\alpha$ , a strain optimized for cloning purpose, and *BL21-DE3*, optimized for protein expression. Cells were spread in LB-agar medium and grown overnight at 37°C. A single colony was then inoculated into 5ml of liquid LB medium and left overnight at 37°C. One ml of culture was then inoculated to 100 ml of 2XYT and grown at 37°C until optical absorbance at 550nm reached a value of 0.6. The growth was then stopped by incubating on ice for 10 minutes. Four 10 ml aliquotes were transferred in 13ml tubes and centrifuged at 5000 rpm for 10 minutes at 4°C. The supernatant was then eliminated and the bacterial pellet resuspended in 1 ml of sterile CaCl<sub>2</sub> 100mM. After 20 minutes of incubation on ice the centrifugation is repeated and the pellet is gently resuspended in 1 ml of sterile CaCl<sub>2</sub> 100mM. Sterilized glycerol to a final concentration of 13% v/v is added to allow long-term storage of cells at -80°C.

### *Transformation*

Chemically competent cells, stored at -80°C, are thawed on ice. Aliquots of 100  $\mu$ l of the bacterial suspension are then added with 2.5  $\mu$ l of ligase product in a sterile 2ml eppendorf tube. After incubating for 30 minutes on ice, transformation was induced by thermic shock, by immersing the tube in a thermostatic bath set at 42°C for 45 seconds and immediately transfer on ice. To allow the rapid growth of transformed cells, 900  $\mu$ l of SOC medium (Table 13) are added to the tube and put in shaking incubator at 37°C for 2 hours. Two aliquots of 100  $\mu$ l and 200  $\mu$ l of bacterial suspension are then plated in

distinct LB-agar Petri dishes added with X-gal, IPTG and the proper antibiotic (Table 13) and incubated at 37°C overnight.

### *Growth medium*

The most common medium for *Escherichia coli* growth is the LB (Luria Bertani, Table 13), while other typologies are mainly used during transformation (SOC, Table 13) or preparation of chemically competent cells (2XYT, Table 13). The preparation is carried out by mixing all the thermostable components in distilled water, adjust pH to neutrality and autoclave sterilizing at 120°C for 20 minutes. Thermolabile compounds like antibiotics and glucose are sterilized by filtration and added after a partial cooling of the autoclaved solution.

In case solid medium was needed (i.e. petri dishes), a quantity of 15g/l agar was added to the solution prior to autoclaving. When blue/white screening was needed to check transformants, 1ml/l IPTG solution (24mg/ml isopropyl  $\beta$ -D-1- thiogalactopyranoside in water) and 1.5 ml/l of X-gal solution (20 mg/ml of 5-bromo-4-chloro-indolyl- $\beta$ -D-galactopyranoside in N,N'-dimethylformamide) were added.

<b>LB medium</b>	
Bacto-Tryptone	10 g/l
Bacto-Yeast extract	5 g/l
NaCl	10 g/l
<b>2XYT medium</b>	
Bacto-Tryptone	16 g/l
Bacto-Yeast extract	10 g/l
NaCl	5 g/l
<b>SOC medium</b>	
Bacto-Tryptone	20 g/l
Bacto-Yeast extract	5 g/l
NaCl	10 mM
KCl	2,5 mM
MgCl <sub>2</sub>	20 mM
Glucose	20 mM

*Table 13. Bacterial grow media*

### *Plasmid extraction*

Singularized cells are grown overnight in 5 ml of LB broth with the proper antibiotics. After centrifuging at 5000 rpm for 5 minutes, the supernatant is removed and the bacterial pellet is gently resuspended in LYR buffer (Table 14) and transferred in a 2ml eppendorf tube. The tube is incubated for 10 minutes at room temperature, then 400  $\mu$ l of alkaline SDS, prepared just before use, is added. The solution is gently mixed by inversion until the liquid became clear, viscous and without clumps.

A volume of 300  $\mu$ l of KOAc solution (see Table 14) is added and the tube is thoroughly mixed, producing a white flocculent material composed by cells debris and chromosomal material. After a 30 minutes incubation on ice the sample is spun in a microcentrifuge at max speed for 5 minutes and the supernatant is recovered in a new tube. Plasmidic DNA is then precipitated by the addition of 1 ml of cold isopropanol and left at room temperature for 5 minutes before being centrifuged at max speed for 10 minutes. The supernatant is then discarded and the pellet is washed with 80% ethanol. DNA is then dried, diluted in 100  $\mu$ l of sterile water and quantified using Nanodrop spectrometer.

<b>LYR buffer</b>	
Glucose	10% w/v
EDTA pH 8	10 mM
Tris-HCl pH 8	25 mM
Lysozyme	2 mg/ml
RNAse A	2 mg/ml
<b>Alkaline SDS solution*</b>	
NaOH	0,1 M
SDS	1%
<b>KOAc solution pH 4.8</b>	
KOAc	147,3 g
Acetic acid	115 ml
Distilled water	up to 500 ml

*Table 14. Plasmid extraction solutions \*To be prepared just before use*

## Expression vector construction

pET28-a plasmid (Figure 15) was chosen as expression vector, to be transformed in the BL21-DE strain in *Escherichia coli*. The vector enables N-terminal tagging using an histidine tag that facilitates the purification of the expressed protein upon induction by IPTG.

The cDNA was retrotranscribed from total White Redhaven RNA treated with DNase and *ccd4* alleles were amplified using primer *ccd4*-5FOR and *ccd4*-CDSrev using the high fidelity Herculase II® polymerase. Purified fragments were then A-tailed, inserted in pGEM-t easy vector and transformed in *Escherichia coli DH5α* strain competent cells. Colonies carrying the functional allele were identified by colony PCR using *ccd4*-5for and *ccd4*-SSRrev primers and visualization in polyacrylamide gel. Colonies carrying the low weight-ssr, hence functional allele were grown overnight in LB and plasmids were extracted. After a preliminary check of plasmid size in agarose, two fragments were then sequenced to check the conformity of the cloned fragment with the deduced coding sequence of peach *ccd4*.

Two oligos with 5' tails for the inclusion of restriction sites at the amplicon ends were developed; for the 5' the restriction enzyme *NdeI* was chosen to allow the cloning of the *ccd4* sequence in frame with the Histidin tag site of pET28-a; an *EcoRI* site was inserted at the 3' extremity. Amplification was carried out using the high fidelity Herculase II® polymerase; 1 µg of PCR product and 1 µg of native pET28 were digested in a 20 µl mix containing 2 µl of 10x Red buffer+ (Fermentas, Lithuania), 2 µl each of *EcoRI* and *NdeI*. The reaction was carried out at 37°C for 3 hours and enzymes were subsequently inactivated at 65°C for 10 minutes. The ligase reaction was carried out on a 30 µl volume adding to the restriction reaction a 10 µl mix containing 1x ligase buffer, 1,5 µl of 10mM ATP and 0,4 µl of T4 DNA ligase (Fermentas, Lithuania). The mix was incubated overnight at 4°C and 5 µl were used for transformation of competent *Escherichia coli* BL21-DE3 cells. A positive colony grown on kanamicine selective LB-agar medium was isolated, the plasmid extracted and sequenced to confirm the correct insertion of gene sequence in frame.

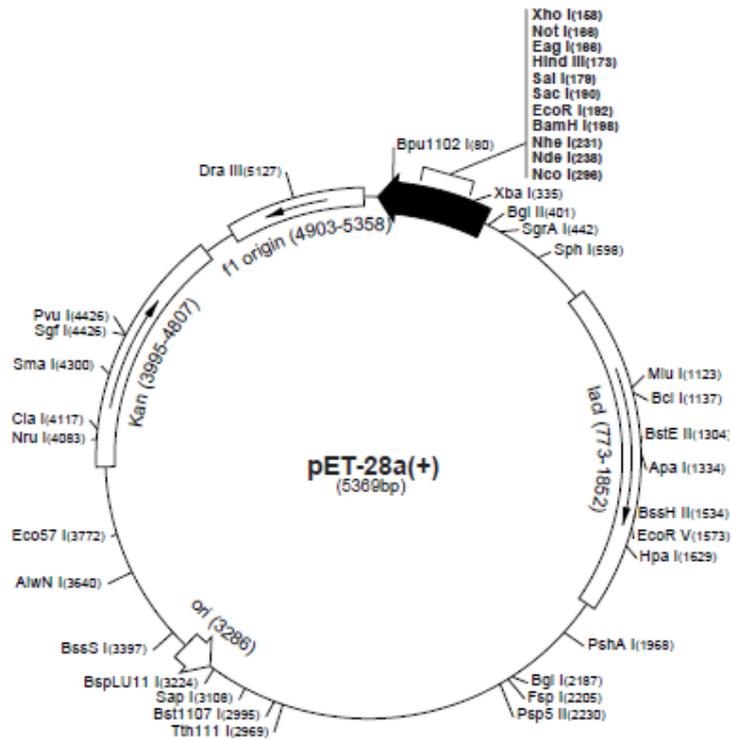


Figure 15. pET-28a native plasmid

## RNA extraction

RNA extraction was carried out according to Zamboni *et al.*, 2008, that adopt a spermidine containing extraction buffer that inactivates RNAses. Nevertheless, all solutions were prepared using DEPC treated water and dry-heat stove treated glassware. Extraction buffer (Table 15) was pre-heated at 65°C in a water bath. Leaves were ground in liquid nitrogen and quickly transferred to the extraction buffer. The sample was vortexed and incubated at 65°C for 15 minutes.

An equal volume of chloroform:isoamyl alcohol 24:1 was then added, and the sample was vigorously vortexed and then centrifuged for 15 min at 13000 g. The supernatant was then carefully transferred to a new tube and the extraction with chloroform:isoamyl alcohol was repeated. The supernatant was then transferred to a new tube and one-third volumes of 8 M LiCl was added. The sample was mixed thoroughly by inverting the tube and incubated overnight at 0 °C (on ice). The sample was then centrifuged at 15500 g for 35 minutes to pellet the RNA and was resuspended in 500 µl of resuspension buffer (Table 15). Samples were re-extracted with an equal volume of chloroform:isoamyl alcohol (24:1 v:v) to reduce residual contaminants; after centrifugation for 10 min at

15000 g the aqueous phase was recovered. Two volumes of ice-cold 100% ethanol were added and RNA was precipitated at -80°C for 30 minutes. The sample was centrifuged at 17000 g for 20 minutes and the supernatant was discarded. The pellet was air-dried at room temperature, resuspended in 20-100 µl of DEPC-treated water.

<b>Extraction buffer</b>	
CTAB	2% w/v
PVP 40000	2% w/vl
Tris-HCl pH 8	100 mM
EDTA	25 mM
NaCl	2 M
Spermidine trihydrochloride	0,05 w/v
β-mercapto-ethanol*	2%
<b>Resuspension buffer</b>	
NaCl	1 M
SDS	0,5 %
Tris-HCl pH 8	10 mM
EDTA	1m M

Table 15. Solutions for RNA extraction according to Zamboni et al. 2008 \*Add just before use

#### *DNase treatment*

The RNA extraction protocol is efficient in removing the most part of DNA. However, to avoid even a minimal genomic DNA contamination, extracted samples were treated with Sigma DNase I kit. The reaction was carried out in a volume of 20 µl. In a RNase free tube, 3 µl of extracted RNA were mixed with 13 µl of DHPC treated water, 2 µl of 10x DNase I buffer and 2 µl of enzyme. The solution was gently mixed and incubated for 15 minutes at room temperature. After incubation time, the enzyme was inactivated by incubating for 10 minutes at 70°C.

#### *Retrotranscription*

To obtain cDNA, total RNA was retrotranscribed using the SuperScript™ II Reverse Transcriptase kit from Invitrogen. Two micrograms of total RNA and 1 µl of oligo-dT (12-18) primer (500 µg/ml) were transferred into a 2ml eppendorf tube together with 9 µl of

DEPC water. After an incubation at 65°C for 5 minutes, the tube was chilled on ice. In the same tube were added 5 µl of 5x reaction buffer, 2 µl of DTT 0,1 M and 1 µl of dNTPs (10mM each); after an accurate resuspension, the mix was incubated at 42°C for 2 minutes. One µl of SuperScript™ II RT enzyme was then added and the reaction was incubated at 42°C for 60 minutes. After that the sample was transferred on ice.

## Mapping

SSR markers pchgms3 and PacA18 (Sosinski *et al.*, 2000; Decroocq *et al.*, 2003) were amplified using the same standard protocol used for the *ccd4*-SSR and visualized in 5% polyacrylamide gel electrophoresis. A three-primer PCR assay was used to map the *ccd4* gene on the F<sub>1</sub> 'Royal Prince' × 'Yoshihime' population (84 individuals). The reaction mixture contained 100 µg genomic DNA, 2.5 mM MgCl<sub>2</sub>, 0.5 mM each dNTP, 0.5 µM *ccd4*-SSRfor and *ccd4*E2rev, 0.08 µM *ccd4*-INS1for primer (see Table 6) and 0.8U Taq polymerase (Amplitaq Gold, Applied Biosystems, Foster City, CA, USA). The following PCR conditions were used: denaturation at 94°C for 7 minutes; 35 cycles of: 94°C for 20 seconds, annealing for 1 minute at 58°C and primer extension for 2 minutes 30 seconds at 72°C; final extension at 72°C for 10 minutes. PCR products were digested with AluI (Fermentas, Hanover, MD, USA) at 37°C for 3 hours, according to manufacturer instruction. Restriction fragments were visualized on 5% polyacrylamide gel. All the markers were mapped using Joinmap 3.0 (Van Ooijen & Voorrips, 2001) using the Kosambi mapping function with a LOD score of 7.

## In-silico analysis

Several bioinformatics programs and tools were used for sequence data analysis, alignment and search in database

### *Sequence manipulation*

- FinchTV 1.4.0 (Geospiza inc. <http://www.geospiza.com>) was used to read and manipulate chromatogram files from the most popular formats
- Ape "A plasmid Editor" 2.0.45 (M. Wayne Davis <http://biologylabs.utah.edu/jorgensen/wayned/ape/>) is the perfect notepad when working with sequences
- Seqman from Lasergene suite v. 7.1.0 (Dnastar inc. <http://www.dnastar.com>) was used for sequence assembly
- Lalign ([http://www.ch.embnet.org/software/LALIGN\\_form.html](http://www.ch.embnet.org/software/LALIGN_form.html)) for local alignment of sequence pairs
- Blast (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) for the retrieval of sequences from online databases
- CENSOR (Kohany *et al.*, 2006) for the screening of repetitive elements in Repbase database (<http://www.girinst.org/censor/index.php>)
- ESTree EST database (<http://www.itb.cnr.it/estree/>)

### *Phylogenetic analysis*

A pool of plant CCDs and NCEDs was selected, including those published by Huang *et al.*, 2009 Ahrazem *et al.*, 2010. This pool was considered representative of plant carotenoid dioxygenases variability, as it includes sequences of the different dioxygenases classes from different taxa.

MEGA 5 was used for multiple sequence alignment, estimation of genetic distances and the creation of a phylogenetic tree. Proteins were aligned by clustalW and the evolutionary history was inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Poisson correction method and statistical support for the topology was obtained by bootstrap analysis with 1000 replicates.

Sequence name	Accession number	Description
AchiCCD7	GU206813.1	Actinidia chinensis carotenoid cleavage dioxygenase 7 (CCD7)
AchiCCD8	GU206812.1	Actinidia chinensis carotenoid cleavage dioxygenase 8 (CCD8)
AthaNCED3	AB026549.1	Arabidopsis thaliana gene for neoxanthin cleavage enzyme
CaraCCD1	DQ157170.1	Coffea arabica carotenoid cleavage dioxygenase 1 (CCD1)
CcleNCED3	DQ309332.1	Citrus clementina 9-cis-epoxycarotenoid dioxygenase 3 (NCED3)
CcleNCED5	DQ309329.1	Citrus clementina 9-cis-epoxycarotenoid dioxygenase 5 (NCED5)
CmorNCED3	AB247159.1	Chrysanthemum x morifolium 9-cis-epoxycarotenoid dioxygenase
CrsatCCD2	AJ132927.1	Crocus sativus mRNA for crocetin dialdehyde
CrsatCCD4	EU523663.1	Crocus sativus chromoplast carotenoid cleavage dioxygenase 4b (CCD4b)
CsinNCED1	DQ028471.1	Citrus sinensis 9-cis-epoxycarotenoid dioxygenase 1
CsinNCED2	DQ028472.1	Citrus sinensis 9-cis-epoxycarotenoid dioxygenase 2
CusatCCD7	HQ005419.1	Cucumis sativus cultivar inbred line 602 carotenoid cleavage dioxygenase 7
DcarNCED1	DQ192200.1	Daucus carota subsp. sativus putative 9-cis epoxycarotenoid dioxygenase (NCED1)
GladNCED4	JF804768.1	Gladiolus hybrid cultivar cultivar Rose Supreme 9-cis-epoxycarotenoid dioxygenase
GlutNCED2	AY466118.1	Gentiana lutea 9-cis-epoxycarotenoid dioxygenase (NCED2)
GmaxCCD7	HM366150.1	Glycine max carotenoid cleavage dioxygenase 7 (CCD7)
GmaxCCD8	HM366151.1	Glycine max carotenoid cleavage dioxygenase 8 (CCD8)
InilNCED1	HQ641566.1	Ipomoea nil 9-cis-epoxycarotenoid dioxygenase 1 (NCED1)
LforNCED3	GQ168942.1	Lilium formosanum NCED3
LspeNCED3	GQ168943.1	Lilium speciosum NCED3
MdomCCD4	EU327777.1	Malus x domestica carotenoid cleavage dioxygenase 4 (CCD4)
MhupNCED2	EU716329.1	Malus hupehensis var mengshanensis 9-cis-epoxycarotenoid dioxygenase
NtabCCD4	JF947192.1	Nicotiana tabacum cultivar Samsun carotenoid cleavage dioxygenase 4 (CCD4)
OfraCCD1	AB526197.1	Osmanthus fragrans CCD1 mRNA for carotenoid cleavage dioxygenase 1
OsatNCED5	AY838901.1	Oryza sativa 9-cis-epoxycarotenoid dioxygenase 5 (NCED5)
PhybCCD1	AY576003.1	Petunia x hybrida carotenoid cleavage dioxygenase 1 (CCD1)
PhybCCD7	FJ790878.1	Petunia x hybrida carotenoid cleavage dioxygenase 7 (CCD7) gene, complete cds
<b>PperCCD4</b>		<b>Prunus persica cvr White Redhaven carotenoid cleavage dioxygenase 4 (CCD4)</b>
PvulNCED1	AF190462.1	Phaseolus vulgaris 9-cis-epoxycarotenoid dioxygenase (NCED1)
RdamCCD4	EU334433.1	Rosa x damascena carotenoid cleavage dioxygenase 4 (CCD4)
SlycCCD7	GQ468555.1	Solanum lycopersicum cultivar M82 carotenoid cleavage dioxygenase 7 (CCD7)
StubNCED1	AY662342.1	Solanum tuberosum 9-cis-epoxy-carotenoid dioxygenase 1
StubNCED2	AY662343.1	Solanum tuberosum 9-cis-epoxy-carotenoid dioxygenase 2
VvinCCD1	AY856353.1	Vitis vinifera 9,10[9',10']carotenoid cleavage dioxygenase (CCD1)
VvinNCED1	AY337613.1	Vitis vinifera 9-cis-epoxycarotenoid dioxygenase 1 (NCED1)
VvinNCED2	AY337614.1	Vitis vinifera 9-cis-epoxycarotenoid dioxygenase 2 (NCED2)

Table 16. List of dioxygenases accession used in phylogenetic analysis

### Rice GAAS analysis

RiceGAAS (<http://ricegaas.dna.affrc.go.jp/>) was used to predict open reading frames within the peach chromosome 1 region corresponding to the Y locus. This service integrates programs for prediction and analysis of protein-coding gene structure. The sequence is analyzed for homology against protein and rice EST databases, gene prediction using various programs as well as analysis of exons, splicing sites, repeats and transfer RNA. All analysis results are then made available using a web-based graphical view.

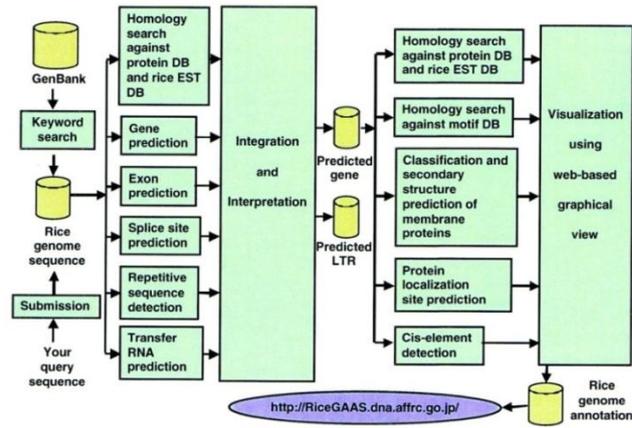


Figure 16. System flowchart of RiceGAAS

The peach genome sequence is available at the following websites:

-[www.peachgenome.org](http://www.peachgenome.org)

-[http://services.appliedgenomics.org/gbrowse/prunus\\_public/](http://services.appliedgenomics.org/gbrowse/prunus_public/)

-[www.phytozome.net/peach](http://www.phytozome.net/peach)

## Results



## Molecular characterization of mutant pairs

To check the genetic identity between Redhaven and White Redhaven, a set of 197 SNPs evenly spread along the genome were used for fingerprinting. The complete absence of polymorphism was confirmed (Table 17) as already reported for 18 SSR markers in Brandi, 2010.

Scaffold	RH		RHB		Tot.
	Homozygous	Heterozygous	Homozygous	Heterozygous	
1	17	8	17	8	25
2	4	14	4	14	18
3	3	30	3	30	33
4	17	27	17	27	44
5	2	11	2	11	13
6	5	18	5	18	23
7	5	12	5	12	17
8	5	19	5	19	24

Table 17 SNPs analyzed to confirm the identical genotypic asset between Redhaven and White Redhaven. The total number of homozygous , heterozygous and total SNP of both accessions per scaffold is reported.

A set of 12 peach microsatellites were used to fingerprint the cultivars, Caldesi 2000 and Cristina, resulting in a complete absence of polymorphisms (Figure 17)

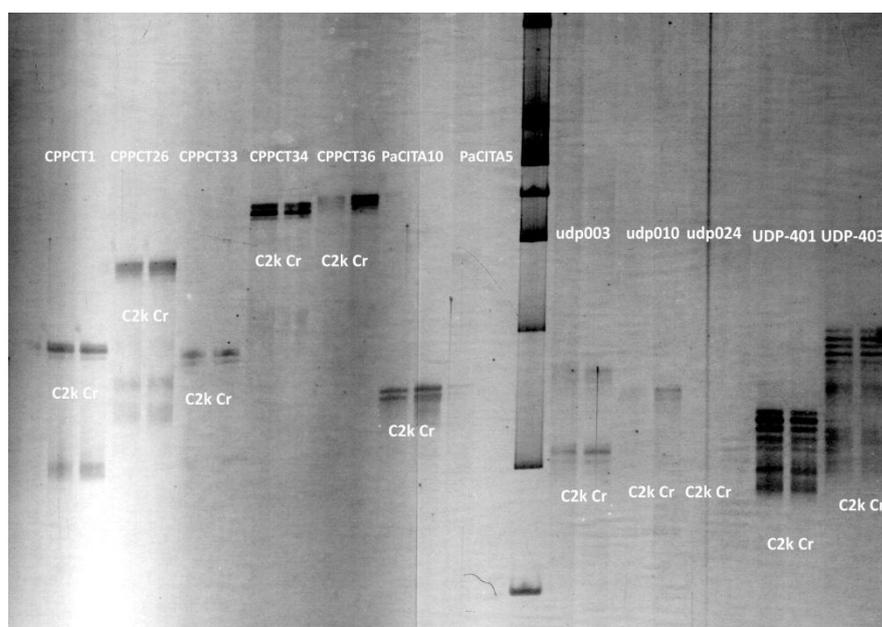


Figure 17. Fingerprinting of cultivars Caldesi 2000 (C2k) and Cristina (Cr) using a set of 12 SSR markers. PaCITA 5 and udp003 showed no signal, while the remaining were identical

## Identification in-silico of genes related to white-yellow phenotype within the Y locus

When this work started, a draft version of the peach genome was available for research groups involved in the International Peach Genome Initiative (IPGI) at [http://services.appliedgenomics.org/gbrowse/prunus\\_public/](http://services.appliedgenomics.org/gbrowse/prunus_public/).

It consisted of unannotated scaffolds, approximately matching peach chromosomes, and the only implemented functions were BLAST and GBrowse; nevertheless, the high quality of this sequence turned out to be a precious tool.

The first step was to try to investigate the sequence defined by the available Y locus linked markers and search for candidate genes. The full SSR sequence of PacA18 and the primers identifying the pchgms3 locus were blasted on the peach genome. As expected both SSRs markers were located in scaffold 1. In detail, the PacA18 SSR was located in the region between 22'795'062 bp and 22'795'743 bp. The pchgms3 reverse and forward primers were located at 27'691'861 bp and 27'692'065 bp, respectively. The two SSRs resulted bracketing a 4'898'906 bp interval on the scaffold 1 (

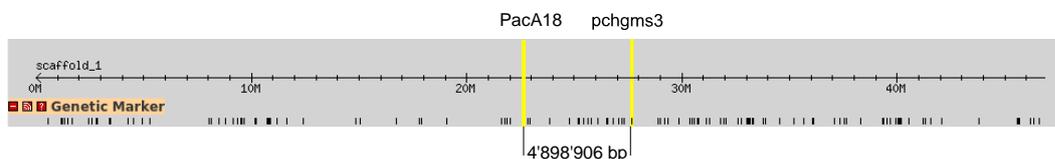


Table 18. SSR markers location on scaffold 1 by GBrowse (IGA Udine website). The Y locus region of 4'898'906 bp is highlighted

This large, unannotated sequence was submitted to RiceGAAS, allowing to predict 1088 putative peptides (see Appendix 1). Among them a putative carotenoid cleavage dioxygenase (*ccd4*) was identified, starting at position 2'846'446 bp while any additional carotenoid-pathway related gene was predicted.

The physical position relative to the Y locus markers was compatible with the available genetic map and the identified putative protein matched several peach EST expressed in mesocarp. This peptide was showing a high similarity with other carotenoid cleavage dioxygenases from *Malus domestica*, *Ricinus communis*, *Citrus clementina*, *Rosa damascena* and *Chrysanthemum morifolium* (Figure 18). Both literature and experimental clues pointed out this putative CCD4 protein as the best candidate gene for controlling fruit flesh color in peach.

<a href="#">Predicted Function</a> [GFS]	<b>by Gene Function Selector (GFSelector)</b>	
	putative carotenoid cleavage dioxygenase 4	
<a href="#">Blastn</a> [db:dna_all]	( <a href="#">DT454560</a> ) PP_YEb0026A05 Peach developing fruit mesocarp Stage S... 1150 0.0	< Alignment >
	( <a href="#">DN677210</a> ) PP_YEa0026A05 Peach developing fruit mesocarp Stage S... 1150 0.0	< Alignment >
	( <a href="#">AM289732</a> ) Prunus persica EST, clone Skin13G12 1116 0.0	< Alignment >
<a href="#">Blastp</a> [db:nr]	gb ABY47995.1  carotenoid cleavage dioxygenase 4 [Malus x domest... 919 0.0	< Alignment >
	ref XP_002519944.1  9-cis-epoxycarotenoid dioxygenase, putative ... 880 0.0	< Alignment >
	ref XP_002268404.1  PREDICTED: hypothetical protein [Vitis vinif... 879 0.0	< Alignment >
	ref XP_002307055.1  predicted protein [Populus trichocarpa] >gi ... 872 0.0	< Alignment >
	ref XP_002326037.1  predicted protein [Populus trichocarpa] >gi ... 867 0.0	< Alignment >
	gb ABC26011.1  carotenoid cleavage dioxygenase 4a [Citrus clemen... 864 0.0	< Alignment >
	gb ABK96278.1  unknown [Populus trichocarpa x Populus deltoides] 862 0.0	< Alignment >
	gb ABY60886.1  carotenoid cleavage dioxygenase 4 [Rosa x damascena] 856 0.0	< Alignment >
	dbj BAE72094.1  Lactuca sativa carotenoid cleavage dioxygenase 1 825 0.0	< Alignment >
	dbj BAF36656.2  putative carotenoid cleavage dioxygenase [Chrysa... 824 0.0	< Alignment >
<a href="#">HMMER</a> [db:Pfam]	<a href="#">RPE65</a> Retinal pigment epithelial membrane protei 777.9 4e-230 1	< Alignment >
	<a href="#">Extensin_2</a> Extensin-like region -170.7 8.9 1	< Alignment >
ProfileScan [db:prosite]	No Hit	-
<a href="#">GO</a> [InterPro]	No GO annotation	-

Figure 18. Analysis summary of the putative *ccd4* gene

## Phylogenetic analysis of carotenoid dioxygenases

A pool of plant carotenoid dioxygenase sequences, including carotenoid cleavage dioxygenases (CCD1, CCD4, CCD7, CCD8) and 9-cis-epoxy-dioxygenase (NCED1, NCED2, NCED3, NCED4, NCED5) was retrieved from online databases (Table 16). The neighbor joining tree (Figure 19) obtained from the protein sequence alignment clearly reflects the distinction between classes of enzymes that share the substrate type but have very different roles in plant metabolism (Auldridge *et al.*, 2006).

For example it is strong the separation between CCD7s and CCD8s, involved in shoot branching signal production (Auldridge *et al.*, 2006) and the other family members. Among them a further division appears between the branch relative to NCEDs, involved in abscisic acid biosynthesis and the CCD1 group that is related to pigment production. Despite the similar enzymatic activity, CCD1 and CCD4 have very distinct cellular localization, the latter being targeted to the plastids by a chloroplast transit peptides that is missing in CCD1 proteins (Ytterberg *et al.*, 2006; Rubio *et al.*, 2008; Walter & Strack, 2011).

A single, highly supported cluster (99% bootstrap) collects all the CCD4 sequences, separated from other enzymes. As expected the peach sequence falls within this cluster, showing the highest sequence similarity with the apple CCD4 (Figure 19).

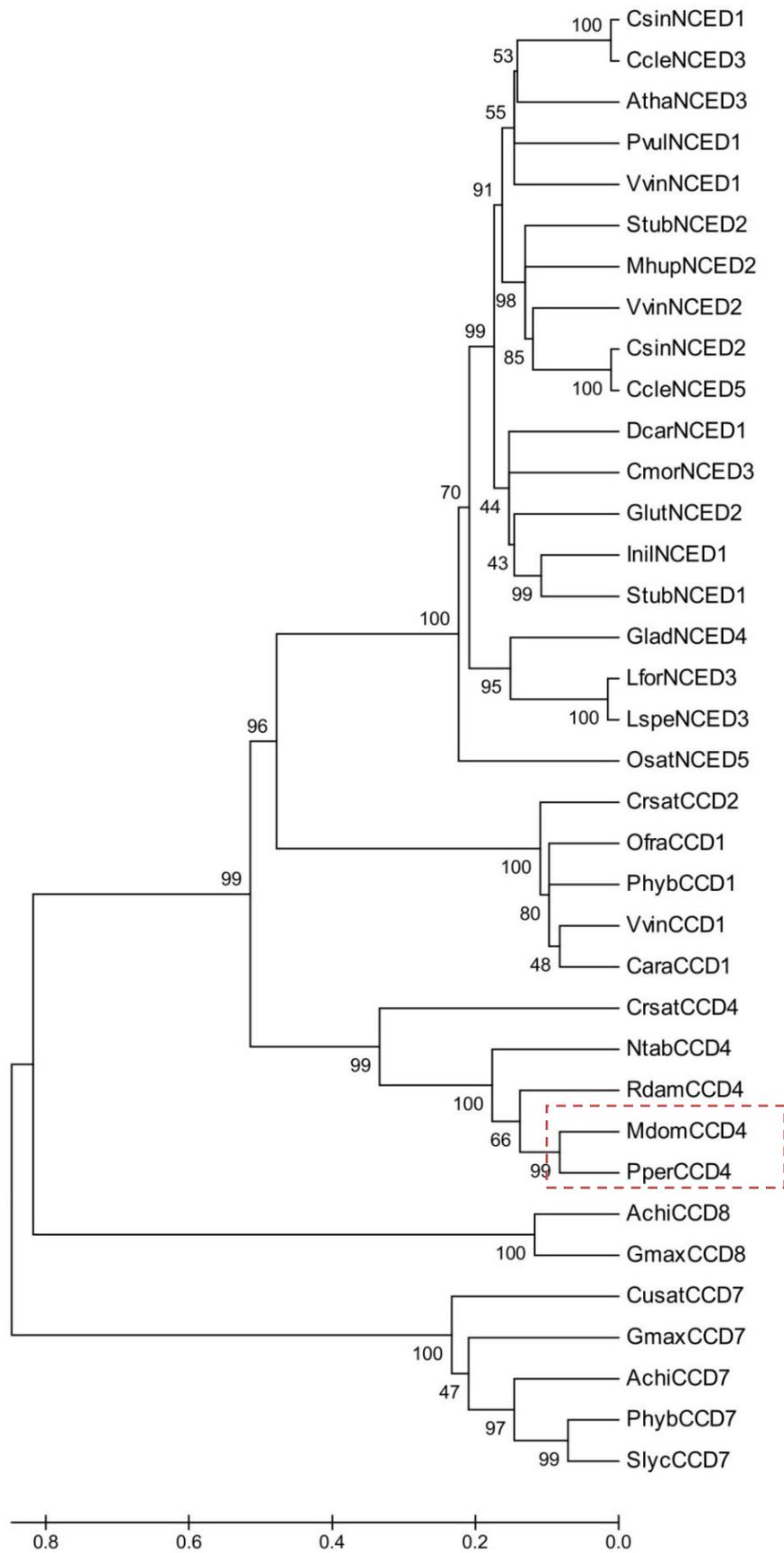


Figure 19. Phylogenetic tree of carotenoid dioxygenase.

## Peach and apple CCD4 comparison

The sequence of the apple *ccd4* mRNA (Acc. # EU327777, Huang *et al.*, 2009) was retrieved from GenBank and aligned on the peach genome v1.0 using BLAST. A single significant hit (E=0) was found on scaffold 1 (from 25639600 to 25641440 bp), showing 85,4% identity at the nucleotide level. This sequence corresponded to the *ccd4* previously predicted by the RiceGAAS analysis. Despite the high homology, the predicted peach protein sequence (ppa006109m) was much shorter than its putative ortholog, counting 427 instead of 591 residues; this difference was due to 136 amino acids missing at the N-terminus in the peach genome annotation of CCD4 (Figure 20).

```

CIUSTAL 2.0.12 multiple sequence alignment

protPP_      IEERPSSPPPASKPTSTKAPQPPKTPSPPLTTKARD--YNNASTFSAAKKGTDPTLPAVI  58
MDeu327777  MEERPSKPPASRPPPSQQP-PRTTTPPLAAKADHALQNNASTFTAAKQ-TVSALPAVI  58
             :*****  ***:*..:: * * .**.***:*** .  *****:***: * .:*****

protPP_      FNALDDIINNFIIDPPLRPSVDPKHVLSSNNFAPVDELPPTECEIIQGSILPPCLDGAYIRNG 118
MDeu327777  FNALDDFINSFIDPPVKPSVDPRHVLSHNFAFVEELPPTECEIIHGSLPPCLDGAYIRNG 118
             *****:*..*****:*****:***:*****:*****:*****:*****

protPP_      PNPQYLPRGPYHLFDGDCMLHSVRISKGRAVLCSRYVKTYKYTIERDAGYPILPSVFSGF  178
MDeu327777  PNPQYLPRGPYHLLDGDGMLHSVRISGGRAVLCSRYVKTYKYTVERDARHPILPNFFSSF  178
             *****:*****:*****:*****:*****:*****:*****:*****

protPP_      NGLTASATRGALSAARVFTGQYNPANGI GLANTSLAFFGNQLYALGESDLPYSLRLTSNG  238
MDeu327777  NGLTASATRGALSAARVLTGQYNPANGI GLANTSLAFFGDRLYALGESDLPYSLRLTSNG  238
             *****:*****:*****:*****:*****:*****:*****:*****

protPP_      DIQTLGRHDFDGKLFMSMTAHPKIDPETGEAFAFRYGFLPPFLTYFRFDANGTKQPDVPI  298
MDeu327777  DIETLGRHDFDGKLSMNMTAHPKIDPDTGEAFAFRYGFI RPFPTYFRFDSNGVKQPDVPI  298
             **:***** * .*****:***** : *****:* .*****

protPP_      FSMVTPSFLHDFAITKKYAI FVDIQIGMNPIDMITKGASPVGLDPSKVPRIGVIPRYAKD  358
MDeu327777  FSMVTPFLHDFAITKKHAI FADIQIGLNLIDMITKRATPFGLDPSKVPRIGVIPRYAKD  358
             *****:*****:***:* ***** * .*****:*****:*****

protPP_      ETEMRFDVPGFNIIHAINAWDEEDAIVM VAPNILSAEHTMERMDLIHASVEKVRIDLKT  418
MDeu327777  ESEMRFVEVPGFNGVHATNAWDEDAIVM VAPNVLSAEHVLERVDLVHCLVEKVRIDLKT  418
             *:*****:***** :* *****:*****:*****:*.**:* . *****

protPP_      GIVSRQPISTRNLDFAVFNPA YVGKKNKYVYAAVGDPMFKISGVVKLDVSNVEHKECIVA  478
MDeu327777  GIVTRQSISTRNLDFAVINPA YLGRKNKYVYAAVGDPMFKISGVVKLDVSNVEHKECIVA  478
             ***:* .*****:***:*. *****:*****:*****:*****:*****

protPP_      SRMFGPGCYGGEPFFVAREPENPEA EDDG YVVTYVHDEKAGESFLVMDAKSPRLDIVA  538
MDeu327777  SRMFGPGCYGGEPFFVAREPENPEA EDNGLVSYVHDEKAGESRFLVMDAKSPQLDIVA  538
             *****:*****:*****:*.**:* ***** *****:*****

protPP_      DVRLPRRVPYGFHGLFVKESDLNKL  563
MDeu327777  AVRMPRRVPYGFHGLFVRESDLNKL  563

             **:*****:*****:***

```

Figure 20. Alignment between peach and apple proteic sequence of putative CCD4. In grey are highlighted the 136 amino acids missing at the N-terminus in the peach genome annotation.

Furthermore, in peach the initial methionine was apparently replaced by an isoleucine. Based on the the fact that the genotype used for genome sequencing was yellow-

fleshed, we firstly hypothesized that a SNP mutation could have replaced the original ATG start codon with an ATT coding for an isoleucine, impairing the functionality of the enzyme. To check this hypothesis, two primers ccd4-5For and ccd4-E2rev flanking the putative mutated start codon were designed on the peach sequence, the fragment was amplified and sequenced in four varieties: the two parents of the segregating population (Royal Prince and Yoshihime) and the two flesh color mutants Redhaven and White Redhaven. The aim was to see if a variant of this sequence with a functional ATG start codon could exist in the white-fleshed phenotype.

After analyzing the sequences of Royal Prince, Yoshihime, Redhaven and White Redhaven they all showed the same ATT codon for isoleucine (see Figure 21). However, a polymorphism on a short microsatellite (slightly imperfect) upstream the presumed start codon was found. At first glance the number of repetitions seemed to be related with the flesh colour: yellow genotypes, like Royal Prince, Redhaven and Lovell, the variety whose genome was sequenced, seemed to possess a TC repetition more than the white variety Yoshihime. All the sequences showed a homozygous pattern, except for White Redhaven that was instead showing an putative heterozygous pattern. This difference between Redhaven and White Redhaven was evinced by the presence in the sequence of the latter ofg double peaks starting just after one of the SSR end (Figure 21). These double peaks could derive from the simultaneous PCR amplification of the two alleles of different size in a heterozygous genotype. Also a sequence retrieved from the ESTree EST database from the white fleshed cultivar Yumyeong was presenting an SSR length of two bp shorter than yellow cultivars .

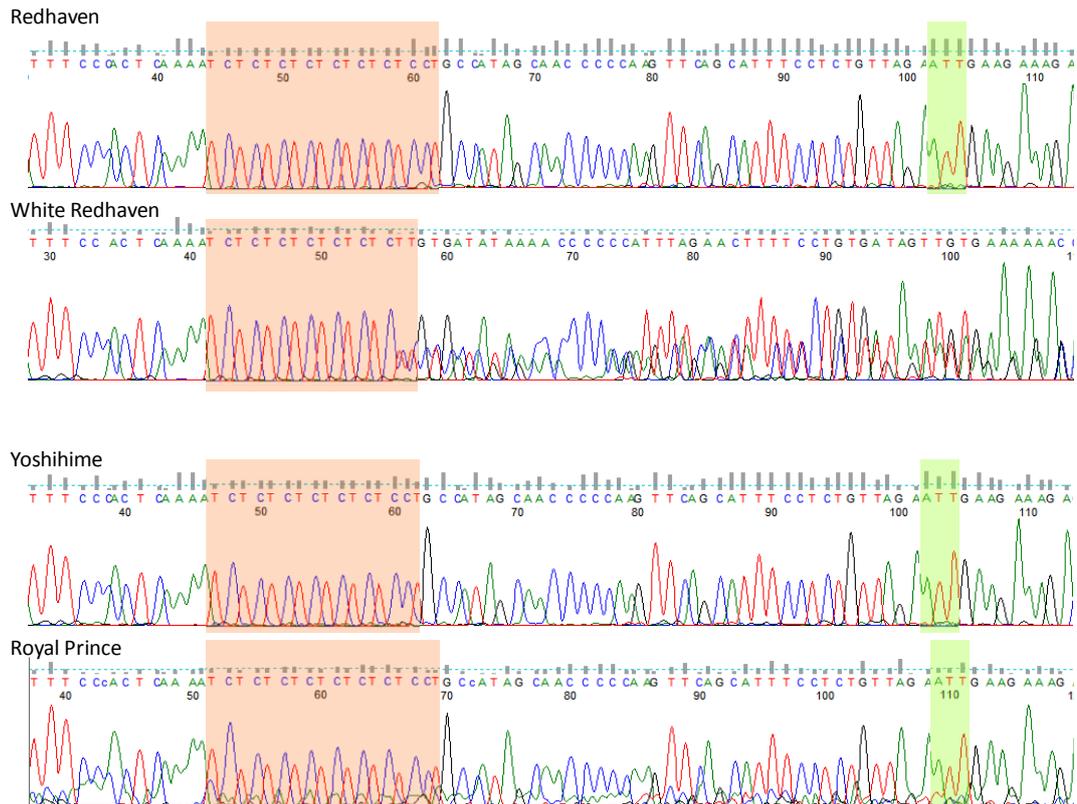


Figure 21. Preliminary sequence of yellow and white varieties in the proximity of the apple ATG codon. In red the microsatellite, in green the ATT codon corresponding to the apple start codon

The microsatellite is not present in the apple sequence, however its site falls within the translated region encoding the predicted chloroplast transit peptide. A two bases addition should result in a frameshift, and would thus affect the prediction of the correct coding sequence (CDS) because of the generation of an aberrant stop codon. When the allele with the (TC)<sub>8</sub> was used for an in silico prediction, its deduced protein sequence counted 597 amino acids including the predicted transit peptide showing a 81,5% identity with the apple protein (Figure 20).

However, the fact that also Yoshihime, known to bear both yellow and white alleles at Y locus, was homozygous for the short (TC)<sub>8</sub> allele seemed to be an inconsistency against the correlation microsatellite length-flesh color. This because the flesh color was found to segregate in the Royal Prince × Yoshihime (Brandi, 2010).

## Development and characterization of ccd4-SSR marker

To analyze more in depth the SSR region in the *ccd4* gene, the *ccd4*-SSRrev primer, was designed to evaluate the microsatellite length in samples deriving from leaf and

fruit flesh of Redhaven and White Redhaven. In the latter, the DNA of the yellow suture was also analyzed. When genomic DNAs were tested, all samples from both genotypes showed the same heterozygous pattern, in contrast with the homozygous status of Redhaven deduced after direct sequencing of the PCR fragment ccd4-5For - ccd4-E2rev (Figure 22). To investigate this discrepancy a two-step PCR protocol was designed. In the first step the genomic DNA was amplified by using the two locus specific primer ccd4-5for and ccd4-E2rev (used in the characterization of the start codon). The amplification product was diluted 100 times and a nested PCR was performed with the two SSR flanking primers (ccd4-5for – ccd4-SSRrev). By using this procedure the homozygous pattern reappeared in the yellow tissues of Redhaven and yellow suture of White Redhaven

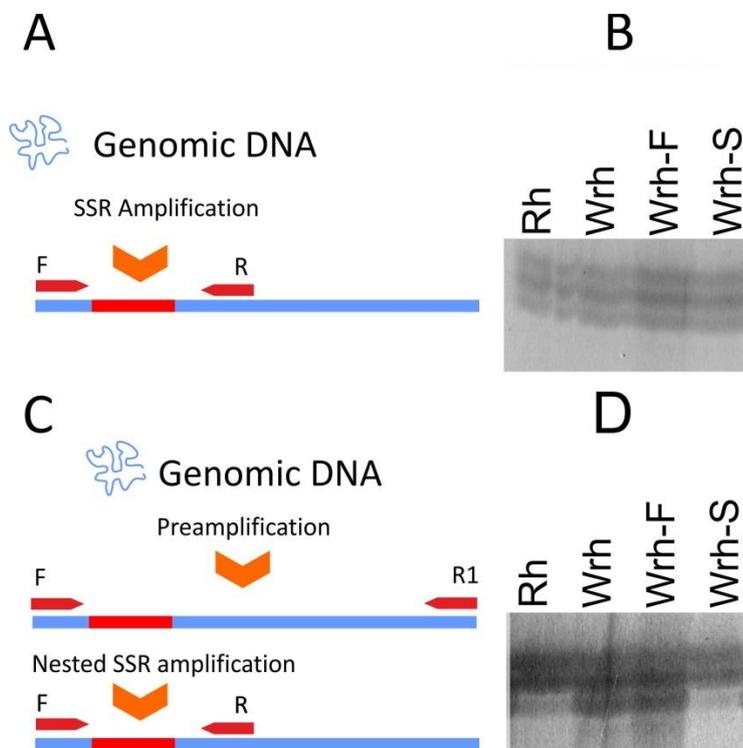


Figure 22. Comparison between ccd4 SSR amplification from genomic template (A) and from PCR nested template (B). The following primers were used: ccd4-5for (F), ccd4-SSRrev (R) and ccd4-E2rev (R1). DNA templates were Redhaven leaf (Rh), White Redhaven leaf (Wrh), White Redhaven flesh (Wrh-F), White Redhaven suture (Wrh-S)

It was thus hypothesized that an unknown mutation could prevent the amplification of the allele bearing the short (TC)<sub>8</sub> microsatellite allele in Redhaven and in the suture of White Redhaven. To fully sequence each allele of the gene, it was necessary to isolate to avoid the problem of direct sequences on total DNAs. To do this, homozygous embryos derived from self-fertilization of Redhaven and White Redhaven were used. Their

genotype was identified using the microsatellite sequence itself (Figure 23). An embryo homozygous for each SSR allele of Redhaven and White Redhaven was selected for directed gene sequencing. This allowed to get around the inability of amplifying, and therefore cloning, the Redhaven allele.

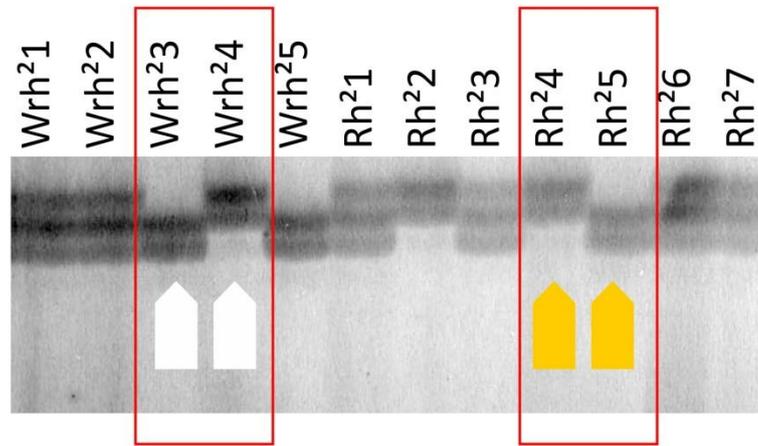


Figure 23. Genotyping of White Redhaven  $\times$  White Redhaven ( $WRh^2n$ ) and Redhaven  $\times$  Redhaven ( $Rh^2n$ ) embryos using *ccd4* ssr.

Several primers (see Table 6) spanning the gene sequence were then designed in order to check the possible existence of a deletion impairing the annealing site of the *ccd4*-E2rev primer and with the intention, at a later stage, to use them to full sequence all the alleles.

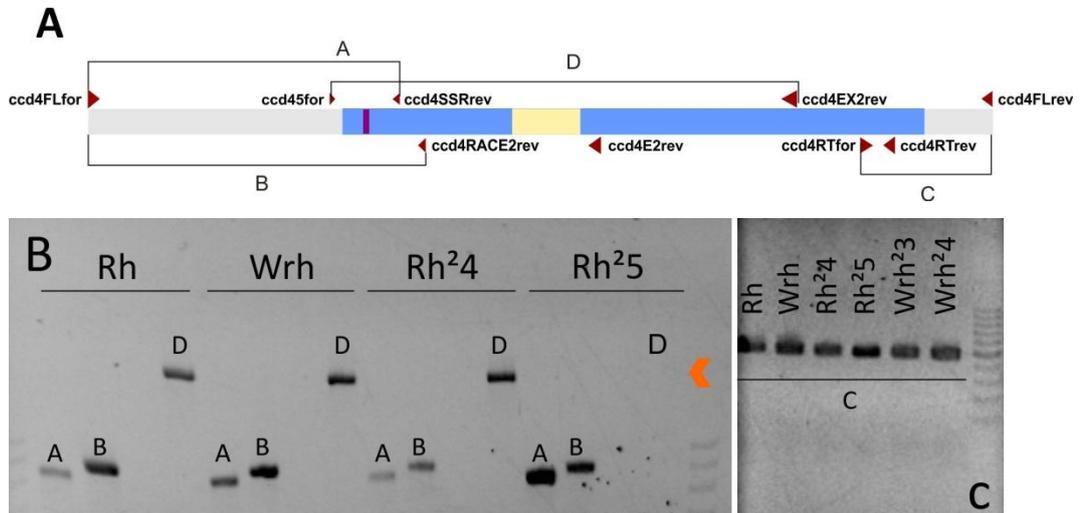


Figure 24. Amplifications of fragments spanning *ccd4* sequence. A: Primers combination used in amplification and gene fragments B: Results of the amplification of different *ccd4* parts in Redhaven, White Redhaven and homozygous embryos, To test this hypothesis, a long-range PCR was performed using Herculase II® on Redhaven, White Redhaven and the homozygous embryos  $Rh^25$  and  $WRh^23$ .

By amplifying several areas of the gene with different primer combinations (Figure 18 A, Figure 24) on Redhaven, White Redhaven and homozygous embryos, it appeared evident that only the fragment D (Figure 24 B), corresponding to the central part of the sequence, was not amplified in the homozygous embryo Rh<sup>25</sup>. It was then hypothesized that lack could be due to the presence of a large insertion within this region in the allele inherited by Redhaven.

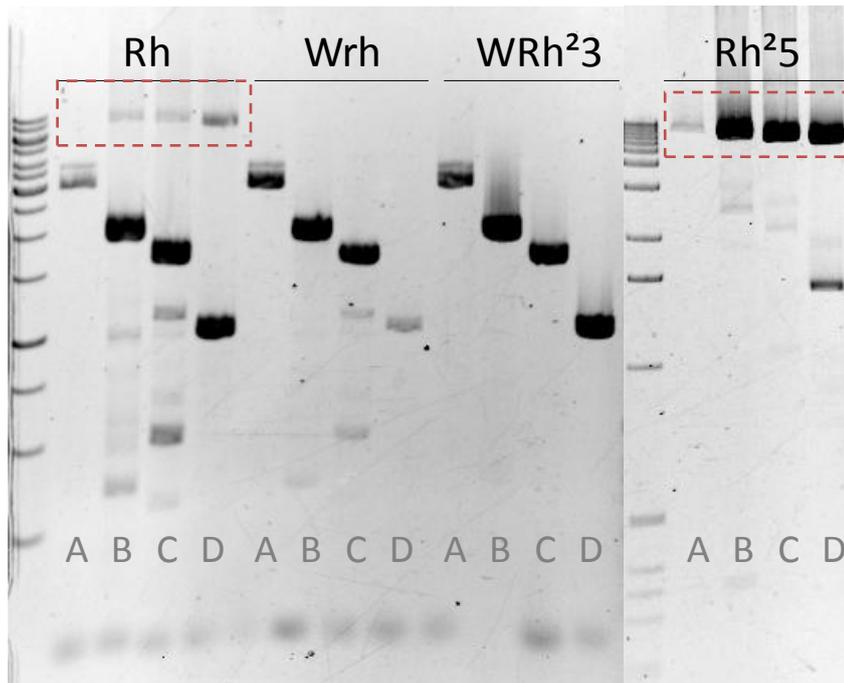


Figure 25. Long range PCR carried on full length alleles of Redhaven (Redhaven), White Redhaven (WRh) and two homozygous embryos derived from White Redhaven (WRh<sup>23</sup>) and from Redhaven (Rh<sup>25</sup>) bearing the (CT)<sub>8</sub> allele using different combinations of primers A: ccd4FLfor - ccd4FLrev; B ccd45for - ccd4RTrev; C ccd45For - ccd4EX2rev; D ccd45for - ccd4E2rev; . The heaviest ladder band corresponds to 10kbp. The dashed boxes highlight the amplification pattern of the heavy allele from Redhaven and Rh<sup>25</sup>

Long range PCR showed that Rh<sup>25</sup> alleles size was several time bigger than the White Redhaven (TC)<sub>8</sub> allele (Figure 25).

## Cloning and characterization of the Redhaven (TC)<sub>8</sub> allele

To investigate further the nature of this insertion, absent in White Redhaven, it was necessary to clone it. To simplify the cloning procedure of a such large fragment and to

avoid sequencing complications due to repetitive sequences, it was decided to subclone the whole Rh<sup>25</sup> allele into smaller fragments.

The restriction enzyme *EcoRI* was chosen its relatively low cutting frequency, due to its 6 bp recognition site and, most important of all, the fact that there weren't cutting sites within the known sequence of the gene. The digestion gave origin to two fragments of about 3000 bp and one of 2500 bp. The next step was the singularization of each fragment. Being sizes so similar, it was not possible to simply cut and subclone all the bands simultaneously. In order to efficiently clone all the fragments, a two-step strategy based on adapter ligation and PCR was optimized. In detail, after the *EcoRI* restriction, the resulting fragments were ligated with an Eco<sub>0</sub> AFLP adapter. In this way each fragment could be selectively amplified using the proper oligo combination in different PCR reactions. The central fragment was amplified using Eco<sub>0</sub> adapter, while the external fragments were amplified using Eco<sub>0</sub> and the proper locus specific primer (ccd4-FLfor for the 5' extremity and ccd4-FLrev for the 3' end). Each band was then cloned in *Escherichia coli* in a pGEM-t easy plasmid and sequenced by primer walking using primers in Table 8.

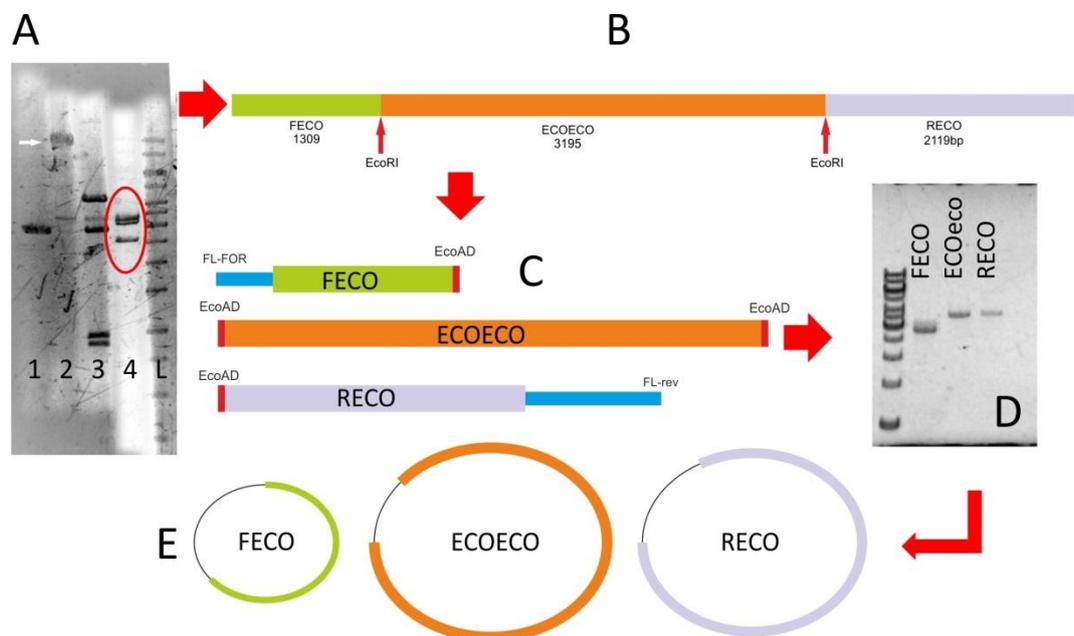


Figure 26. Subcloning of digested band – A: 1 Undigested insertionless full length *ccd4* allele from *Wrh<sup>23</sup>*; 2 Undigested fragment with insertion from *Rh<sup>25</sup>*; 3 fragment restriction using *PstI*; 4 fragment restriction using *EcoRI*; B Scheme of the restricted fragment; C Restriction fragments after adapret ligation; D Selective amplification of digested fragment (lane A4); E Plasmids obtained after cloning

The three sequences from the (TC)<sub>8</sub> allele of Redhaven were assembled and led to the identification of a 6263-bp insertion within the intron, with two 490bp direct repeats at

the extremities. The Censor analysis of insertion sequence pointed out analogies with *Gypsy* and *Copia*-like LTR retrotransposon. A CATA insertion site was found at both ends of the inserted sequence. The correspondent (TC)<sub>8</sub> allele of White Redhaven didn't show any trace of insertion and a single CATA sequence was found. Therefore, the mutation associated with the phenotype reversion to white most likely consisted in the excision of the transposable element, and can be considered a retro-mutation of the (TC)<sub>8</sub> allele (Figure 36 and Figure 27).

Collectively, during the analysis of Redhaven and White Redhaven three distinct alleles were characterized. Since the white flesh phenotype is dominant over yellow, the two alleles from Redhaven were putatively associated with the yellow flesh color, and they were tentatively named  $y^1$  (for the (TC)<sub>9</sub> allele) and  $y^2$  (for the (TC)<sub>8</sub> allele with the large LTR insertion). The allele  $y^1$  resulted present in White Redhaven as well; consistently with the heterozygous status of this mutant for the flesh color, its second allele, identical to  $y^2$  but lacking the insertion, was putatively associated with the white phenotype and named  $W^1$  (Figure 27). Three sequences were submitted to GeneBank with the accession numbers JX309999 ( $W^1$ ), JX310000 ( $y^1$ ), JX310001 ( $y^2$ ).

All the alleles have two exons and a single intron. Prediction of open reading frame on these sequences highlighted that the alleles  $W^1$  and  $y^2$  possess a full 1794 bp ORF, encoding a protein of 597 residues. In the  $y^1$  allele, the additional repeat in the microsatellite region causes a frameshift after 24 codons and a premature stop at position 72 from ATG (Figure 27).

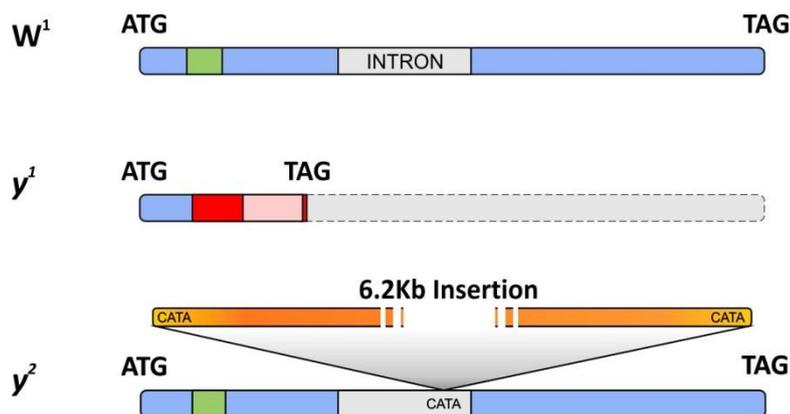


Figure 27. Structure of the functional ( $W^1$ ) and mutated ( $y^1$ ) *ccd4* alleles. The coding sequence (light blue) contains a 209 bp intron (grey) and a STR consisting of 8 repetitions of the dinucleotide (TC) (green). In  $y^1$ , the STR consists of 9 (TC) repetitions (red), resulting in a shift of the reading frame (pink) and a premature end of translation.  $y^2$  mutation has a 6.8-kb insertion (orange) in correspondence of a CATA site 38bp before the 3' end of the intron.

## Analysis of the chimerical fruit tissues of White Redhaven

Tissues extracted from suture and flesh of White Redhaven chimeric fruit were analyzed and tested to determine their genotype using the three primer system (see page 40). As expected, the pattern of the tissue was coherent with the predicted color: the white flesh shows the presence of  $W^1$  allele, while the yellow suture of the same fruit shows the presence of  $y^2$ , like the non-mutated Redhaven.

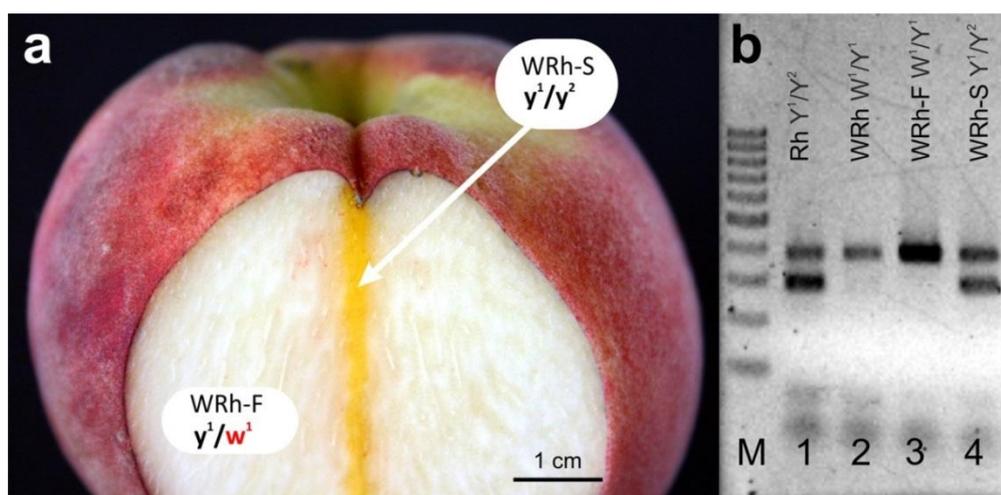


Figure 28. Chimeric status of White Redhaven fruit. (A) The section in correspondence of the fruit suture reveals the yellow sector maintaining the ancestral genotype; (B) Analysis of different tissue samples from White Redhaven fruit using the three primer system: 1-2 DNA extracted from leaf; 3 Amplification of DNA extracted from White Redhaven flesh doesn't show  $y^1$ ; 4 Amplification of White Redhaven DNA extracted from the yellow tissue suture shows the same pattern as Redhaven. M: 100bp ladder (Fermentas, Lithuania)

## Characterization of Caldesi 2000/Cristina mutant system

The second mutant system analyzed consisted of the white-fleshed cultivar Caldesi 2000 and its yellow-fleshed mutant Cristina. The SSR analysis directly distinguishes the two clones: Caldesi 2000 resulted heterozygous for the microsatellite length, while Cristina resulted homozygous for the  $(TC)_9$  allele, as shown in Figure 29. The analysis with three primer system excluded the presence of the LTR insertion in any  $ccd4$  allele in both cultivars. Therefore, Caldesi 2000 has a solid heterozygous  $W^1/y^1$  genotype while in the yellow fruit tissues of Cristina, only the  $y^1$  allele is present. Thus the insertion of two bases turns the functional  $W^1$  allele of Caldesi 2000 into the  $y^1$  allele of Cristina (Figure 29 B and C).

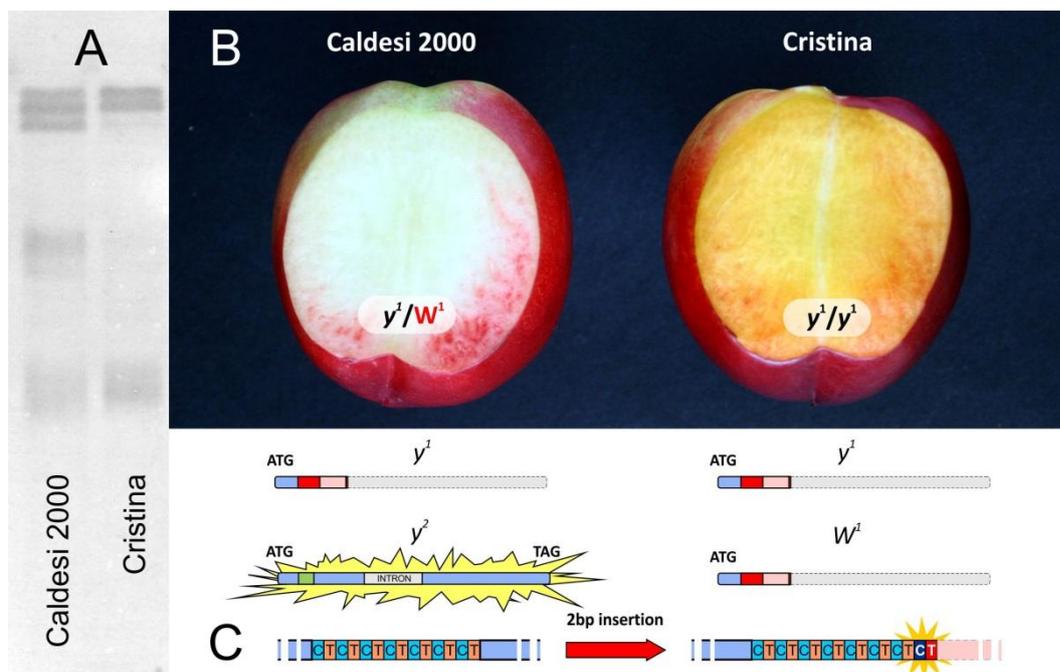


Figure 29. Characterization of the Caldesi 2000 to Cristina mutation. (A) *ccd4* SSR polymorphism (B) Fruit section and allelic composition; (C) The mechanism of mutation that involves the elongation of the *ccd4* microsatellite that causes the inactivation of the  $W^1$  functional allele of Caldesi 2000.

## Mapping and validation of *ccd4* co-segregation with Y-locus

The *ccd4* mapping in the Royal Prince  $\times$  Yoshihime progeny resulted very complicated. The molecular analysis showed that both genotypes are heterozygous, being Royal Prince  $y^1/y^2$  and Yoshihime  $W^1/y^2$ . The  $y^2$  allele is common between the two parents, therefore the expected segregation type is "ABxAC". Unfortunately, neither the *ccd4* SSR segregation, neither the retrotransposon insertion detected with the three primer method, are sufficiently precise in identifying simultaneously the different alleles. The *ccd4* SSR segregate as "ab x aa" in the progeny, being Royal Prince heterozygous and Yoshihime homozygous (Figure 30). Therefore the *ccd4* can be mapped only in Royal Prince.

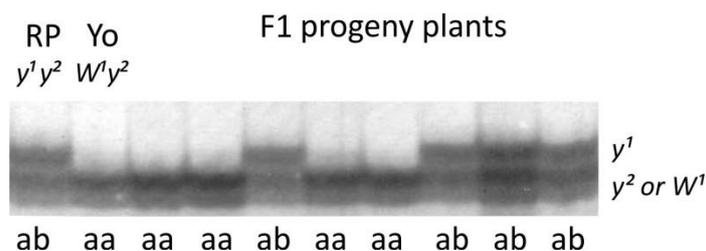


Figure 30. Segregation of *ccda4* SSR marker in RP  $\times$  Yo

The LTR insertion is present in both genotypes and its segregation type is "AB x AB" (Figure 31). The expected segregation (1:2:1) could be used for mapping in both parents but the correlation with the fruit flesh color is hampered by the fact that it is not possible to distinguish the  $W^1$  and  $y^1$  alleles. For that half of the heterozygous seedlings are yellow and half are white. Therefore, the linkage with the flesh color cannot be established.

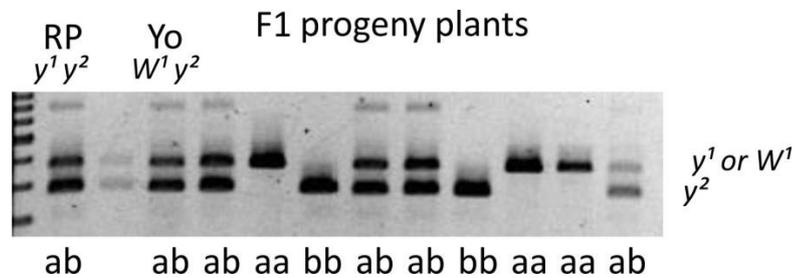


Figure 31. Segregation of insertion marker in  $Rp \times Yo$

For that reason a strategy to simultaneously detect the SSR polymorphism and the insertion was designed (see page 55). The results of this analysis are summarized in figure z. The progeny had 41 white and 43 yellow-fleshed individuals, consistent with the 1:1 ratio expected for a Mendelian trait in a cross between a heterozygous and a recessive homozygous genotype.

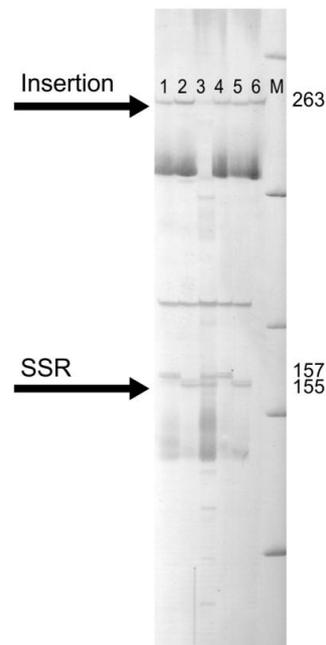


Figure 32. Segregation of  $ccd4$  alleles in the Royal Prince  $\times$  Yoshihime progeny. After amplification with primers  $ccd4$ -SSRfor,  $ccd4$ -E2rev and  $ccd4$ -INS1for (see Table 6) followed by digestion with  $AluI$ , a 263bp band denotes the presence of the  $y^2$  allele, whereas 155 and 157bp fragments identify  $W^1$  and  $y^1$ , respectively.

The presence of PacA18 and pchgms3 markers confirm the position of the *Y* locus on constructed linkage group 1. In fact it was placed between markers PacA18 (at 10.6 cM distance) and pchgms3 (at 7.1 cM). As expected, *ccd4* gene fully co-segregated with flesh color trait and the *W<sup>1</sup>* allele from Yoshihime was associated with white flesh phenotype (Figure 33). The position of *Y* locus and *ccd4* was also verified on the peach whole genome sequence and PacA18 was found at 22795743 bp, *Y* locus/*ccd4* at 25639436 bp and pchgms3 at 27692065 bp.

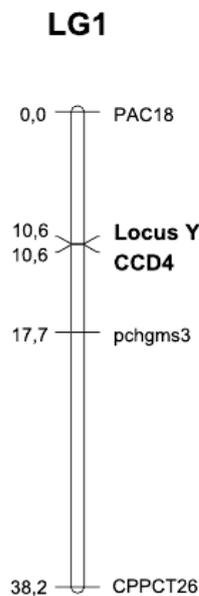


Figure 33. Frame of the RP x Yh map showing the co-localization of *Y* locus and *ccd4*

## Analysis of *ccd4* in peach germplasm

Specific PCR markers were then designed to rapidly assay the presence of the insertion and to estimate the repetitive sequence length (not in a single PCR run). These markers were then used to characterize a set of peach genotypes. These included two pairs of mutant sports, from white to yellow and vice-versa, more than 106 different genotypes representative of the peach germplasm variability and including white and yellow fleshed genotypes and a yellow/white segregating progeny.

This panel of genotype was including some of the most important American founders widely used in peach breeding (). Based on the three alleles characterized in the two

mutant systems, all cultivars bearing at least one copy of  $W^1$  were predicted to be white-fleshed; this prediction was consistent with the observed phenotype in all cases except 'Gialla Tardiva', 'Jonia' and 'OroA'. Despite being yellow-fleshed, these cultivars initially seemed to possess the allele  $W^1$ . However, sequencing of *ccd4* from these cultivars highlighted the presence of a previously unidentified mutation, named  $y^3$  (Acc. # KC142158). It consists of a single base substitution (A to T) at position 1520 of the coding sequence that causes a codon change from lysine (AAG) to a stop codon (TAG), resulting in a truncated protein lacking 91 C-terminal residues (

Figure 34). A specific TSP-SNP marker was thus developed to assay the distribution of this allele in the complete set of cultivars; however, none of the remaining cultivars was possessing  $y^3$  besides the three aforementioned.



Figure 34. Structure of the  $y^3$  allele found in peach germplasm  $y^3$  mutation consists of a A/T substitution at position 1520 of the coding sequence that originates a premature stop codon, resulting in a truncated protein that lacks 91 C-terminal residues.

Finally, all the 47 white-fleshed varieties carried the  $W^1$  allele, in homozygosis (9 cultivars) or coupled with an  $y^n$  allele (33 with  $y^1$  and 5 with  $y^2$ ); conversely, the 59 yellow-fleshed cultivars consistently had two mutated alleles (Table 19.) as expected.

	Cultivar	Flesh Color	Y locus	Mutations <sup>a</sup>			ccd4 Genotype	
				y <sup>1</sup>	y <sup>2</sup>	y <sup>3</sup>		
	Adriana	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Alired	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Alitop	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Alix	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Ambersisters	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Andross	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Azurite	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Babygold 9	Y	y/y	-/-	+/+	-/-	y <sup>2</sup>	/ y <sup>2</sup>
	Big Top	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Copia Poa	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Coraline	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Cristina <sup>b</sup>	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Diamond Ray	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Egea	Y	y/y	-/-	+/+	-/-	y <sup>2</sup>	/ y <sup>2</sup>
	Eolia	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Guglielmina	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Honey Blaze	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Honey Kist	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Jonia	Y	y/y	+/-	-/-	+/-	y <sup>1</sup>	/ y <sup>3</sup>
	Jungerman	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Kaweah	Y	y/y	+/-	+/-	-/-	y <sup>1</sup>	/ y <sup>2</sup>
	Lady Erica	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Maria Aurelia	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Maria Dolce	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Maria Dorata	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Max 7	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Maycrest	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Nectaross	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	O'Henry	Y	y/y	+/-	+/-	-/-	y <sup>1</sup>	/ y <sup>2</sup>
	Oro A	Y	y/y	+/-	-/-	+/-	y <sup>1</sup>	/ y <sup>3</sup>
	Red Top	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Red Valley	Y	y/y	+/-	+/-	-/-	y <sup>1</sup>	/ y <sup>2</sup>
	Redhaven	Y	y/y	+/-	+/-	-/-	y <sup>1</sup>	/ y <sup>2</sup>
	Rich Lady	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Rome Star	Y	y/y	+/-	+/-	-/-	y <sup>1</sup>	/ y <sup>2</sup>
	Rose Diamond	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Royal Glory	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Royal Prince	Y	y/y	+/-	+/-	-/-	y <sup>1</sup>	/ y <sup>2</sup>
	Royal Summer	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
	Rubirich	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>

Worldwide released varieties

Worldwide  
released  
varieties

September Free	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/	y <sup>1</sup>
Valley Red	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/	y <sup>1</sup>
Velvetsisters	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/	y <sup>1</sup>
Vistarich	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/	y <sup>1</sup>
Zee Lady	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/	y <sup>1</sup>
Alba	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Aliblanca	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Alipersiè	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Amanda	W	W/W	-/-	-/-	-/-	W <sup>1</sup>	/	W <sup>1</sup>
Artic Sweet	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Benedicte	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Caldesi 2000	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Crizia	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Douceur	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Early Giant	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Early Silver	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Emeraude	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Gladis	W	W/y	-/-	+/-	-/-	W <sup>1</sup>	/	y <sup>2</sup>
Greta	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Jade	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Kurakatawase	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Maria Anna	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Maria Delizia	W	W/W	-/-	-/-	-/-	W <sup>1</sup>	/	W <sup>1</sup>
Maylis	W	W/W	-/-	-/-	-/-	W <sup>1</sup>	/	W <sup>1</sup>
Neve	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Pearlsisters D93 1-19	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Platifortwo	W	W/W	-/-	-/-	-/-	W <sup>1</sup>	/	W <sup>1</sup>
Sahong	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
September Snow	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Silver Giant	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Silver Rome	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Snow Queen	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Spring Snow	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Spring White <sup>c</sup>	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Stark Saturn	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Summer Sweet	W	W/y	-/-	+/-	-/-	W <sup>1</sup>	/	y <sup>2</sup>
Tendresse	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Vanilia	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
White Redhaven <sup>d</sup>	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>
Yoshihime	W	W/y	-/-	+/-	-/-	W <sup>1</sup>	/	y <sup>2</sup>
Yumyeung	W	W/W	-/-	-/-	-/-	W <sup>1</sup>	/	W <sup>1</sup>
Zephir	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/	y <sup>1</sup>

CRA	FRF								
Advanced Selections		IFF 331	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/ y <sup>1</sup>
		IFF 813	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
Heritage Italian Germoplasm		Fuoco di Romagna	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Gialla Tardiva	Y	y/y	-/-	-/-	+/+	y <sup>3</sup>	/ y <sup>3</sup>
		Percoca di Romagna	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Percoca di Romagna 7	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Pesca Carota	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Bella di Cesena	W	W/W	-/-	-/-	-/-	W <sup>1</sup>	/ W <sup>1</sup>
		Bella di Cesena precoce	W	W/W	-/-	-/-	-/-	W <sup>1</sup>	/ W <sup>1</sup>
		Bella di Piangipane	W	W/W	+/+	-/-	-/-	W <sup>1</sup>	/ W <sup>1</sup>
		Buco Incavato	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/ y <sup>1</sup>
		Iris Rosso	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/ y <sup>1</sup>
		Rosa del West	W	W/y	+/-	-/-	-/-	W <sup>1</sup>	/ y <sup>1</sup>
	S.Anna Balducci	W	W/W	-/-	-/-	-/-	W <sup>1</sup>	/ W <sup>1</sup>	
US breeding founders		Admiral Dewey	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Early Crawford	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Elberta	Y	y/y	+/-	+/-	-/-	y <sup>1</sup>	/ y <sup>2</sup>
		J.H.Hale	Y	y/y	+/-	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Kalamazoo	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Muir	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Rio Oso Gem	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Yellow St. John	Y	y/y	+/+	-/-	-/-	y <sup>1</sup>	/ y <sup>1</sup>
		Chinese Cling	W	W/y	-/-	+/-	-/-	W <sup>1</sup>	/ y <sup>2</sup>
	Georgia Belle	W	W/y	-/-	+/-	-/-	W <sup>1</sup>	/ y <sup>2</sup>	

	W <sup>1</sup>	y <sup>1</sup>	y <sup>2</sup>	y <sup>3</sup>
W <sup>1</sup>	9	33	5	0
y <sup>1</sup>	-	47	7	2
y <sup>2</sup>	-	-	2	0
y <sup>3</sup>	-	-	-	1
Color	47	59		
	white	yellow		
Total	106			

Table 19. *ccd4* genotype classes frequency the analyzed peach accessions; see **Errore. L'origine riferimento non è stata trovata.** for the list of analyzed genotypes.

## Expression and functional verification

A pET28 plasmid containing the functional Pp-ccd4-W<sup>1</sup> allele was built using the cDNA derived from White Redhaven. It is a vector optimized for protein expression and purification. The plasmid (Figure 35. pET28-ccd4 vector map) contains a 1752 bp ORF encoding a protein of 583 aminoacids, corresponding to the full 563 aa CCD4 sequence plus a 20 aa N-terminal Histidine Tag for protein purification. The ORF is under control of the LacZ promoter and its expression can be induced by the addition of IPTG.

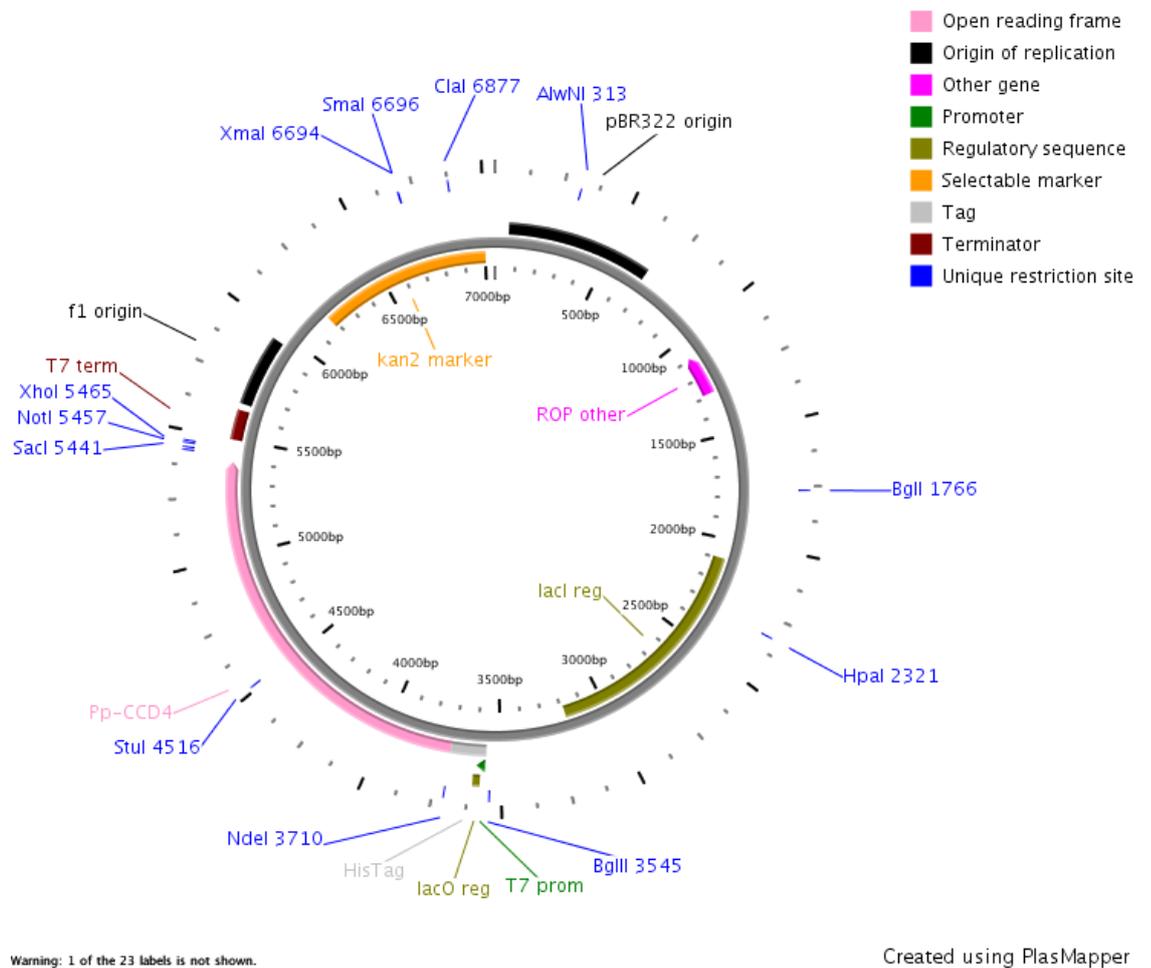


Figure 35. pET28-ccd4 vector map



## Discussion

In this work a *ccd4* gene was found in the *Y* locus of peach; the annotation was given on the base of a high sequence homology with the apple *ccd4* sequence. Of course functional verification of the role of this gene is needed; vectors for heterologous expression in bacterial cells were produced, which will allow the purification of the recombinant protein and biochemical analysis aimed to confirm its activity as CCD4. This annotation is strongly supported by a phylogenetic analysis in which in a pool of plant dioxygenases, PpCCD4 was placed in the CCD4 cluster with a strong statistical support.

In this cluster a panel of CCD and NCED sequences in GeneBank were analyzed and results clearly confirm the primary annotation of the genemade on the basis of the homology between the apple and the peach. The high homology of this gene and of the upper part of the peach linkage group 1 was finally in a recent work. The comparison of the three available rosaceae genomes, peach, diploid strawberry and apple allowed to identify an elevated conservation of synteny between the apple chromosomes 13 and 16 and the upper part of peach chromosome, that harbours the *Y* locus.

A first validation of the role of *ccd4* as gene responsible for the flesh color change in peach was deduced after a characterization of different *ccd4* alleles in a pool of peach accessions including two flesh color mutants in which the mutation of *ccd4* is associated to the color change. In White Redhaven the excision of a transposable element interrupting the gene sequence restored the gene functionality and its ability to promote carotenoid degradation, resulting in a change of the flesh color from yellow to white. Conversely, in Cristina, a microsatellite length variation created a frame shift in the functional allele, inactivating the gene and causing the phenotypic shift to yellow.

In both cases, as the mutation involved a single bud layer (L2), tissues originated by other layers retained the ancestral genotype, as confirmed by the molecular markers developed on those mutations. The original flesh color phenotype is still visible in these chimeric fruits in correspondence of the suture, that is originated by layer L1 (Dermen & Stewart, 1973).

Summarizing the results obtained by analyzing a pool of peach varieties, the presence of a functional *ccd4* allele (*W*<sup>1</sup>) always associated with the white color of the flesh. Conversely, the yellow phenotype is visible only in cultivars bearing other alleles.

Mapping activity and *in silico* analysis demonstrated that the *ccd4* gene is located in correspondence of the *Y* locus. In fact the segregation of a *ccd4*-specific marker in a progeny of 84 individuals from the cross Royal Prince (yellow) × Yoshihime (white) evidenced that *ccd4* and *Y* locus are co-mapping. Even in this case the functional  $W^1$  allele is fully associated to the white flesh phenotype, which is visible only in those seedlings that inherited the  $W^1$  allele from Yoshihime.

Three the different mechanisms of mutation which altered the functionality of this gene were identified.

The  $W^1$  and  $y^1$  alleles differ only in the number of repeats in the microsatellite region; since this site falls within the ORF, the additional dinucleotide in the  $y^1$  allele causes a frameshift, resulting in an aberrant CDS and a truncated protein. It is likely that a spontaneous mutation in the microsatellite length of the  $W^1$  allele occurred in the L2 layer, producing a mutated allele that is identical to  $y^1$  and causing the flesh color change from white to yellow (Figure 29. Characterization of the Caldesi 2000 to Cristina mutation.). The most common cause of length changes in short sequence repeats is replication slippage: during DNA replication, slipping of DNA polymerase III on the template strand can cause insertion or deletions of repeats in the synthesized strand (Wang *et al.*, 2009). In most cases, the slippage results in a change of just one repeat unit. When the microsatellite is located within protein-coding regions, the length variation can lead to a loss of gene function via frameshift mutation or by interfering with the processing of the primary transcript. Ideally, such a mechanism can also account for a gain of function mutation, by restoring the original microsatellite length, or by adding a number of repeats appropriate to recover the reading frame. Slip-strand mispairing errors are corrected by mismatch repair (MMR) systems, thus SSR stability depends upon a balance between the DNA slippage and the repair effectiveness; when the MMR system is altered, SSR instability increases (Li *et al.*, 2004). Interestingly, Caldesi 2000 is known to be characterized by a high frequency of variation in phenotypic traits, and Cristina is one of its sports (A. Liverani, unpublished data). Analysis of the MMR system elements could be one way to investigate the high variability of this cultivar.

The analysis showed that the  $y^2$  allele from Redhaven and the  $W^1$  allele from White Redhaven differ for the presence of a large transposable element in the intron sequence. In a previous study, the *ccd4* expression level in fruit flesh were shown to be much lower in Redhaven than in White Redhaven (Brandi *et al.*, 2011); this phenomenon could be originated by the large retrotransposon insertion in Redhaven, which could

interferes with expression of the gene, even if it does not alter the coding sequence. In European pear, the transcription of an S-allele is affected by regulatory sequences found in a mobile element located within the intron of the gene itself (Sanzol, 2009). These elements can hamper the synthesis and processing of a functional mRNA by incorporating additional splicing sites or generating chimerical transcripts, which in turn may produce aberrant proteins or be targeted to the mRNA-degradation machinery of the cell (Sanzol, 2009). Retrotransposon insertions are generally stable, as replication occurs through a RNA intermediate, but some cases of excision have been reported. In *Petunia*, the *Hose in Hose* floral homeotic mutant is caused by a retrotransposon insertion in the promoter of *PvGlo* gene, resulting in the conversion of sepals to petals. Subsequent excision of this retrotransposon, associated with epigenetic changes at the locus, caused reversion to normal phenotype and restored wild-type flower development (Li *et al.*, 2010). In this case however, excision of the retrotransposon left a 293-bp footprint; in the case of White Redhaven, on the contrary, the transposable element seems to have undergone a precise excision, without leaving any signature sequences; thus, the excision site results indistinguishable from insertion-less alleles. Similar events have been seldom described in eukaryotic and prokaryotic systems (Kuzin *et al.*, 1994 in *Drosophila* and Nefedova *et al.*, 2006 in *Escherichia coli*).

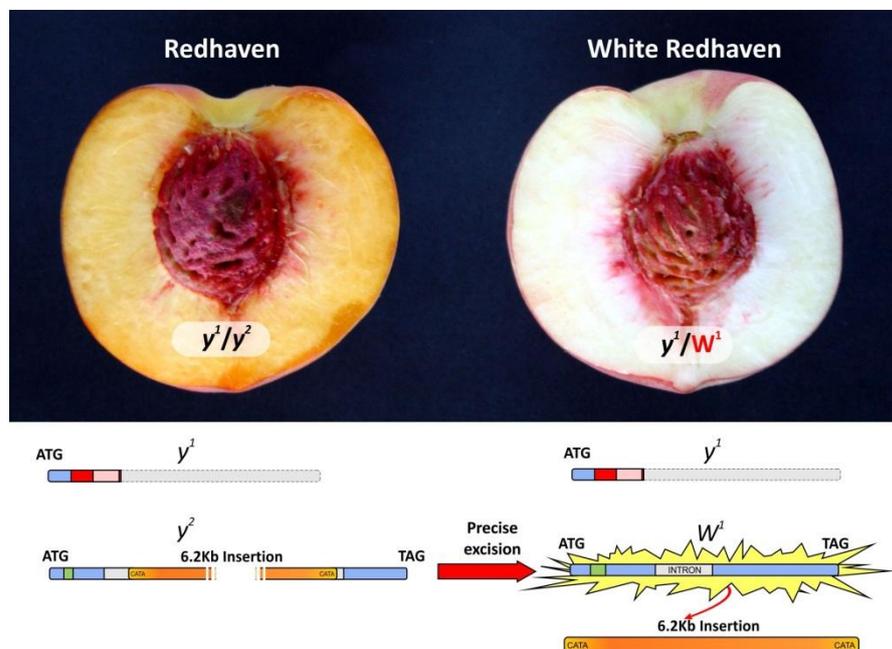


Figure 36 Reversion of  $y^2$  mutation in Redhaven to White Redhaven. Redhaven has a solid  $y^1/y^2$  genotype. In white tissue of White Redhaven fruit flesh, the 6.8-kb insertion is excided and the allele reverts to fully functional  $W^1$ , re-establishing the white phenotype. The CATA signature is duplicated at the 5' and 3' end, but the precise excision of the retrotransposon restores the original single CATA in the resulting allele

A third mutation, ( $y^3$ ) was found in a small number of varieties during the extensive analysis of the pool of peach accessions. This consist in a single nucleotide substitution that generates a premature stop codon at position 1520. When the deduced protein is aligned and confronted with CCD4, it is possible to see that this truncated protein is lacking a consistent number of highly conserved residues, and is thus either inactive or targeted for degradation.

Interestingly, the three observed mutations were due to three different sources of variation in plant genomes. Of these, the mutation observed in allele  $y^3$ , can be considered the most stable one, as the allele functionality can theoretically be restored only by a second mutation affecting the same position and reverting to the original codon, that is extremely unlikely to occur. On the other hand, microsatellite repeats and transposable elements can account for relatively high mutation rates. This can provide an explanation for the occasional observation of phenotype shifts in the vegetative propagation of cultivars, as observed in the cases of Caldesi 2000/Cristina and Redhaven/White Redhaven.

Collectively the findings that yellow phenotype is originated after mutation is in agreement with the hypothesis that the white color is ancestral in peach, and that independent mutations caused flesh color change to yellow during the breeding history of this species. This gave origin to different recessive alleles that, when in homozygosis, bring the yellow color. In this work we characterized four allelic variants, three of which are not functional; the reduced number of alleles can depend on the fact that the American and European peach germplasm has a very narrow genetic base because of the bottleneck caused by the early breeding programs, that relied on a small number of original founder varieties (Aranzana *et al.*, 2010). However it can not be excluded that additional alleles could exist in peach germplasm from the eastern center of diversification of this species.



## Conclusions

Collectively, our results clearly point out *ccd4* as the gene controlling flesh color in peach. Three distinct mutations accounted for the occurrence of yellow phenotype within all the analyzed cultivars; however, the presence of additional *ccd4* mutations in other peach varieties not analyzed in this study cannot be excluded. Moreover, the sequence polymorphisms found within two mutant systems allowed to design specific markers able to discriminate the mutated sports from their standard cultivars that were not distinguishable using markers like SSRs or SNPs. The study of spontaneous mutants, including chimeric genotypes, represents the most reliable approach to prove gene function in this species, in which a functional demonstration cannot be obtained by genetic transformation, due to the woody habit and the recalcitrant behavior of cultured explants.

The knowledge of the genetic bases of the flesh color trait in peach will have an immediate practical outcome, especially in breeding programs. The molecular markers developed in this study enable a fast and accurate Y-locus genotyping of cultivars to be used as parents in programmed crosses and thus the prediction of the segregation of the flesh color trait in their progenies; moreover, the same markers will be a valuable tool towards the early selection of seedlings by MAS (marker assisted selection) that will increase the efficiency of peach breeding programs.

In the future, a functional validation of the annotation of this gene will be performed by heterologous expression of peach CCD4 in bacterial cells and an *in vitro* biochemical characterization of its substrates and products. As efficient regeneration protocols are not yet available in this species, a direct *in vivo* validation of the function of this gene by transformation seems at present hardly feasible; nevertheless, different approaches will be evaluated to test the *in vivo* activity of this enzyme, such as transient silencing of the gene by RNA interference or VIGS.

## **Additional material**



## Appendix 1 - RiceGAAS Predicted function

Candidate gene is found on page 103

pac-pchgms_001	01	putative homeodomain-leucine zipper transcription factor TaHDZip1-2
pac-pchgms_001	02	hypothetical protein
pac-pchgms_001	03	putative hAT family dimerisation domain containing protein
pac-pchgms_001	04	putative calmodulin binding / transcription regulator
pac-pchgms_001	05	hypothetical protein similar to Os02g0509600
pac-pchgms_001	06	putative NTF2B (NUCLEAR TRANSPORT FACTOR 2B); Ran GTPase binding / protein transporter
pac-pchgms_001	07	putative histidine kinase 3
pac-pchgms_001	08	hypothetical protein similar to PREDICTED: proteasome (prosome, macropain) subunit,
pac-pchgms_001	09	hypothetical protein
pac-pchgms_001	10	hypothetical protein
pac-pchgms_001	11	hypothetical protein
pac-pchgms_001	12	putative heat shock protein
pac-pchgms_001	13	putative kinesin family protein
pac-pchgms_001	14	hypothetical protein
pac-pchgms_001	15	hypothetical protein
pac-pchgms_001	16	putative membrane protein
pac-pchgms_001	17	putative Fbox protein
pac-pchgms_001	18	putative ARSK1 (root-specific kinase 1); kinase
pac-pchgms_001	19	putative acyl:coa ligase
pac-pchgms_001	20	putative RecName: Full=Squamosa promoter-binding-like protein 12
pac-pchgms_001	21	putative binding / zinc ion binding
pac-pchgms_001	22	putative plant-specific domain TIGR01627 family protein
pac-pchgms_001	23	putative AAA ATPase
pac-pchgms_001	24	putative ninein isoform 1
pac-pchgms_001	25	putative cysteine desulfurase
pac-pchgms_001	26	putative Aspartate/tyrosine/aromatic aminotransferase
pac-pchgms_001	27	putative oxidoreductase, 2OG-Fe(II) oxygenase family protein
pac-pchgms_001	28	putative RecName: Full=Chlorophyll a-b binding protein 7, chloroplastic;
pac-pchgms_001	29	hypothetical protein
pac-pchgms_001	30	putative zinc finger (MYND type) family protein / F-box family protein
pac-pchgms_001	31	putative RecName: Full=NADH-ubiquinone oxidoreductase 11 kDa subunit;
pac-pchgms_001	32	putative rubber elongation factor (REF) family protein
pac-pchgms_001	33	putative ASY' (ASYNAPTIC 1); DNA binding
pac-pchgms_001	34	hypothetical protein
pac-pchgms_001	35	putative NLI interacting factor (NIF) family protein
pac-pchgms_001	36	hypothetical protein
pac-pchgms_001	37	putative cytochrome P450
pac-pchgms_001	38	hypothetical protein
pac-pchgms_001	39	putative RecName: Full=Mitochondrial import receptor subunit TOM20;
pac-pchgms_001	40	unknown protein
pac-pchgms_001	41	putative Protein kinase domain containing protein
pac-pchgms_001	42	hypothetical protein
pac-pchgms_001	43	putative flagellar inner arm dynein 1 heavy chain beta
pac-pchgms_001	44	putative polyprotein
pac-pchgms_001	45	putative transposon protein Pong sub-class
pac-pchgms_001	46	putative Ulp1-like peptidase
pac-pchgms_001	47	putative IQ motif containing GTPase activating protein 2

pac-pchgms\_001 48 putative HsdR family type I site-specific deoxyribonuclease  
 pac-pchgms\_001 49 putative transposase  
 pac-pchgms\_001 50 hypothetical protein  
 pac-pchgms\_001 51 putative NF-YC13 (NUCLEAR FACTOR Y, SUBUNIT C13); DNA binding / transcription factor  
 pac-pchgms\_001 52 putative MAP65-8 (MICROTUBULE-ASSOCIATED PROTEIN 65-8)  
 pac-pchgms\_002 01 putative ELMO domain-containing protein 2  
 pac-pchgms\_002 02 putative cell division protein FtsZ  
 pac-pchgms\_002 03 putative anthocyanin-O-methyltransferase  
 pac-pchgms\_002 04 hypothetical protein  
 pac-pchgms\_002 05 putative anthocyanin-O-methyltransferase  
 pac-pchgms\_002 06 unknown protein  
 pac-pchgms\_002 07 putative sulfite oxidase  
 pac-pchgms\_002 08 putative exostosin family protein  
 pac-pchgms\_002 09 putative peptidase  
 pac-pchgms\_002 10 hypothetical protein  
 pac-pchgms\_002 11 putative NBS type disease resistance protein  
 pac-pchgms\_002 12 hypothetical protein  
 pac-pchgms\_002 13 putative FAR1; Zinc finger, SWIM-type  
 pac-pchgms\_002 14 putative Viral A-type inclusion protein repeat  
 pac-pchgms\_002 15 putative F-box domain containing protein  
 pac-pchgms\_002 16 putative hAT family dimerisation domain containing protein, expressed  
 pac-pchgms\_002 17 putative polyprotein  
 pac-pchgms\_002 18 putative pentatricopeptide (PPR) repeat-containing protein  
 pac-pchgms\_002 19 hypothetical protein  
 pac-pchgms\_002 20 putative TNP1  
 pac-pchgms\_002 21 hypothetical protein  
 pac-pchgms\_002 22 putative thioredoxin H  
 pac-pchgms\_002 23 putative jacalin lectin family protein  
 pac-pchgms\_002 24 hypothetical protein  
 pac-pchgms\_002 25 putative F-box domain containing protein  
 pac-pchgms\_002 26 putative F-box domain containing protein  
 pac-pchgms\_002 27 putative F-box domain containing protein  
 pac-pchgms\_002 28 putative transposon protein  
 pac-pchgms\_002 29 hypothetical protein  
 pac-pchgms\_002 30 putative F-box family protein  
 pac-pchgms\_002 31 putative 60S ribosomal protein L17 (RPL17B)  
 pac-pchgms\_002 32 putative emb1688 (embryo defective 1688); GTP binding / GTPase  
 pac-pchgms\_002 33 putative transposase  
 pac-pchgms\_002 34 hypothetical protein  
 pac-pchgms\_002 35 putative hAT family dimerisation domain containing protein  
 pac-pchgms\_002 36 putative f-box family protein  
 pac-pchgms\_002 37 putative Nucleoporin p54  
 pac-pchgms\_002 38 putative GCS1 (GLUCOSIDASE 1); alpha-glucosidase  
 pac-pchgms\_002 39 putative gag/pol polyprotein  
 pac-pchgms\_002 40 putative polyprotein  
 pac-pchgms\_002 41 putative EIP28  
 pac-pchgms\_002 42 putative NAD-dependent epimerase/dehydratase  
 pac-pchgms\_002 43 hypothetical protein  
 pac-pchgms\_002 44 putative EMB3011 (embryo defective 3011); ATP binding / RNA helicase/ helicase/ nucleic acid binding  
 pac-pchgms\_002 45 hypothetical protein  
 pac-pchgms\_002 46 putative leucine-rich repeat family protein

pac-pchgms\_002 47 putative 20S proteasome beta subunit E  
 pac-pchgms\_002 48 putative polyprotein  
 pac-pchgms\_002 49 putative polyprotein  
 pac-pchgms\_002 50 hypothetical protein  
 pac-pchgms\_002 51 hypothetical protein  
 pac-pchgms\_002 52 putative armadillo/beta-catenin repeat family protein / U-box domain-containing family protein  
 pac-pchgms\_002 53 putative AAA-type ATPase family protein  
 pac-pchgms\_002 54 putative cytochrome P450  
 pac-pchgms\_002 55 putative Ulp1-like peptidase  
 Pac-pchgms\_003 01 putative lipoprotein  
 Pac-pchgms\_003 02 putative FAR1; Zinc finger, SWIM-type  
 Pac-pchgms\_003 03 hypothetical protein  
 Pac-pchgms\_003 04 putative penicillin-binding protein  
 Pac-pchgms\_003 05 unknown protein  
 Pac-pchgms\_003 06 putative TNP2  
 Pac-pchgms\_003 07 putative Retrotransposon gag protein  
 Pac-pchgms\_003 08 hypothetical protein  
 Pac-pchgms\_003 09 putative En/Spm-like transposon protein  
 Pac-pchgms\_003 10 putative TNP1  
 Pac-pchgms\_003 11 hypothetical protein  
 Pac-pchgms\_003 12 putative TNP2  
 Pac-pchgms\_003 13 putative fusion  
 Pac-pchgms\_003 14 putative cytochrome P450  
 Pac-pchgms\_003 15 hypothetical protein  
 Pac-pchgms\_003 16 putative Yop effector YopM  
 Pac-pchgms\_003 17 putative golgi microtubule-associated protein, isoform B  
 Pac-pchgms\_003 18 putative hAT dimerisation domain-containing protein  
 Pac-pchgms\_003 19 putative cytochrome P450  
 Pac-pchgms\_003 20 unknown protein  
 Pac-pchgms\_003 21 putative Mutator-like transposase  
 Pac-pchgms\_003 22 unknown protein  
 Pac-pchgms\_003 23 putative cytochrome P450  
 Pac-pchgms\_003 24 hypothetical protein  
 Pac-pchgms\_003 25 hypothetical protein  
 Pac-pchgms\_003 26 unknown protein  
 Pac-pchgms\_003 27 putative TNP2  
 Pac-pchgms\_003 28 putative zinc finger, CHC2-type  
 Pac-pchgms\_003 29 putative tryptophan 2-monooxygenase  
 Pac-pchgms\_003 30 hypothetical protein similar to SJCHGC09187 protein  
 Pac-pchgms\_003 31 putative polypeptide with a gag-like domain  
 Pac-pchgms\_003 32 putative TNP1  
 Pac-pchgms\_003 33 putative TNP2  
 Pac-pchgms\_003 34 hypothetical protein  
 Pac-pchgms\_003 35 putative polyprotein  
 Pac-pchgms\_003 36 hypothetical protein  
 Pac-pchgms\_003 37 putative cytochrome P450  
 Pac-pchgms\_003 38 hypothetical protein  
 Pac-pchgms\_003 39 hypothetical protein  
 Pac-pchgms\_003 40 putative TNP2  
 Pac-pchgms\_003 41 putative TNP1  
 Pac-pchgms\_003 42 hypothetical protein  
 Pac-pchgms\_003 43 putative D-alanyl-alanine synthetase A

Pac-pchgms\_003 44 putative cytochrome P450  
 Pac-pchgms\_003 45 putative cytochrome P450  
 Pac-pchgms\_003 46 putative pol-polyprotein  
 Pac-pchgms\_003 47 putative M18 protein precursor  
 Pac-pchgms\_003 48 putative cytochrome P450  
 Pac-pchgms\_003 49 putative Signal transduction histidine kinase regulating C4-dicarboxylate transport system  
 Pac-pchgms\_003 50 putative multicopper oxidase type 2  
 Pac-pchgms\_003 51 hypothetical protein  
 Pac-pchgms\_003 52 hypothetical protein  
 Pac-pchgms\_003 53 hypothetical protein  
 Pac-pchgms\_003 54 putative Cyclin-like F-box; FAR1; Zinc finger, SWIM-type  
 Pac-pchgms\_003 55 putative polyprotein  
 Pac-pchgms\_004 01 putative polyprotein  
 Pac-pchgms\_004 02 hypothetical protein  
 Pac-pchgms\_004 03 hypothetical protein  
 Pac-pchgms\_004 04 putative TNP1  
 Pac-pchgms\_004 05 putative TNP2  
 Pac-pchgms\_004 06 putative TNP2  
 Pac-pchgms\_004 07 putative NAC domain protein  
 Pac-pchgms\_004 08 putative cytochrome P450  
 Pac-pchgms\_004 09 putative polyprotein  
 Pac-pchgms\_004 10 putative cytochrome P450  
 Pac-pchgms\_004 11 putative amelogenin  
 Pac-pchgms\_004 12 putative TNP2  
 Pac-pchgms\_004 13 hypothetical protein  
 Pac-pchgms\_004 14 putative cytochrome P450  
 Pac-pchgms\_004 15 putative polyprotein  
 Pac-pchgms\_004 16 hypothetical protein  
 Pac-pchgms\_004 17 putative golgi microtubule-associated protein, isoform B  
 Pac-pchgms\_004 18 hypothetical protein  
 Pac-pchgms\_004 19 putative nodulin family protein  
 Pac-pchgms\_004 20 putative transposase  
 Pac-pchgms\_004 21 hypothetical protein  
 Pac-pchgms\_004 22 hypothetical protein  
 Pac-pchgms\_004 23 putative Zinc finger, CCHC-type  
 Pac-pchgms\_004 24 putative Retrotransposon gag protein  
 Pac-pchgms\_004 25 hypothetical protein  
 Pac-pchgms\_004 26 putative TNP2  
 Pac-pchgms\_004 27 putative Retrotransposon gag protein  
 Pac-pchgms\_004 28 putative glucose-methanol-choline oxidoreductase  
 Pac-pchgms\_004 29 hypothetical protein  
 Pac-pchgms\_004 30 putative golgi microtubule-associated protein, isoform B  
 Pac-pchgms\_004 31 putative cytochrome P450  
 Pac-pchgms\_004 32 putative Mutator-like transposase  
 Pac-pchgms\_004 33 hypothetical protein  
 Pac-pchgms\_004 34 putative transposon protein Pong sub-class  
 Pac-pchgms\_004 35 putative cytochrome P450  
 Pac-pchgms\_004 36 putative polyprotein  
 Pac-pchgms\_004 37 putative OTU-like cysteine protease family protein  
 Pac-pchgms\_004 38 putative C2 calcium/lipid-binding region-containing protein  
 Pac-pchgms\_004 39 hypothetical protein  
 Pac-pchgms\_004 40 putative NADH dehydrogenase

Pac-pchgms\_004 41 hypothetical protein  
 Pac-pchgms\_004 42 hypothetical protein  
 Pac-pchgms\_004 43 hypothetical protein  
 Pac-pchgms\_004 44 putative protein binding  
 Pac-pchgms\_004 45 putative Helicase associated domain family protein, expressed  
 Pac-pchgms\_004 46 putative gag-pol polyprotein  
 Pac-pchgms\_004 47 putative Helicase associated domain family protein, expressed  
 Pac-pchgms\_004 48 hypothetical protein  
 Pac-pchgms\_004 49 putative urease  
 Pac-pchgms\_004 50 putative lipoxygenase  
 Pac-pchgms\_004 51 putative acyl:coa ligase  
 Pac-pchgms\_004 52 putative metastasis associated 1, isoform CRA\_a  
 Pac-pchgms\_004 53 putative MURF1  
 Pac-pchgms\_004 54 hypothetical protein  
 Pac-pchgms\_004 55 putative protein kinase family protein  
 Pac-pchgms\_004 56 hypothetical protein  
 Pac-pchgms\_004 57 hypothetical protein  
 Pac-pchgms\_004 58 putative trichohyalin  
 Pac-pchgms\_004 59 putative Terpenoid cylases/protein prenyltransferase alpha-alpha toroid; Bacterial adhesion  
 Pac-pchgms\_004 60 putative remorin family protein  
 Pac-pchgms\_004 61 putative emb2421 (embryo defective 2421); monooxygenase/ oxidoreductase  
 Pac-pchgms\_004 62 putative integral membrane protein  
 Pac-pchgms\_005 01 putative helicase  
 Pac-pchgms\_005 02 putative En/Spm-like transposon protein  
 Pac-pchgms\_005 03 putative nucleoside-triphosphatase  
 Pac-pchgms\_005 04 putative iojap family protein  
 Pac-pchgms\_005 05 putative Ketoacyl-ACP Reductase (KAR)  
 Pac-pchgms\_005 06 putative lysine/histidine transporter  
 Pac-pchgms\_005 07 putative ankyrin repeat family protein  
 Pac-pchgms\_005 08 putative TPR domain protein  
 Pac-pchgms\_005 09 putative glyceraldehyde 3-phosphate dehydrogenase  
 Pac-pchgms\_005 10 putative B12D protein  
 Pac-pchgms\_005 11 putative transferase family protein  
 Pac-pchgms\_005 12 hypothetical protein  
 Pac-pchgms\_005 13 putative transferase family protein  
 Pac-pchgms\_005 14 putative protein phosphatase 2A regulatory subunit B'  
 Pac-pchgms\_005 15 putative protein phosphatase 2A regulatory subunit B'  
 Pac-pchgms\_005 16 putative retrotransposon protein  
 Pac-pchgms\_005 17 putative 7S RNA binding  
 Pac-pchgms\_005 18 hypothetical protein  
 Pac-pchgms\_005 19 putative bromodomain containing protein  
 Pac-pchgms\_005 20 putative F-box family protein  
 Pac-pchgms\_005 21 putative DNA binding / nuclease  
 Pac-pchgms\_005 22 putative GT-1  
 Pac-pchgms\_005 23 putative cell cycle checkpoint protein MAD2 homolog  
 Pac-pchgms\_005 24 putative cc-nbs-lrr resistance protein  
 Pac-pchgms\_005 25 hypothetical protein  
 Pac-pchgms\_005 26 putative multidrug resistance protein ABC transporter family  
 Pac-pchgms\_005 27 putative MAD2  
 Pac-pchgms\_005 28 putative CEP350 protein  
 Pac-pchgms\_005 29 hypothetical protein  
 Pac-pchgms\_005 30 putative TNP2

Pac-pchgms_005	31	hypothetical protein
Pac-pchgms_005	32	hypothetical protein similar to carbon monoxide dehydrogenase accessory protein
Pac-pchgms_005	33	putative multidrug resistance protein ABC transporter family
Pac-pchgms_005	34	putative transposon protein
Pac-pchgms_005	35	putative helicase SWR1
Pac-pchgms_005	36	hypothetical protein
Pac-pchgms_005	37	hypothetical protein
Pac-pchgms_005	38	hypothetical protein
Pac-pchgms_005	39	putative transferase family protein
Pac-pchgms_005	40	putative myosin 29
Pac-pchgms_005	41	putative pol-polyprotein
Pac-pchgms_005	42	putative multidrug resistance protein ABC transporter family
Pac-pchgms_005	43	putative copia-like polyprotein
Pac-pchgms_005	44	putative GAG-POL precursor
Pac-pchgms_005	45	putative lysyl-tRNA synthetase
Pac-pchgms_005	46	putative transferase family protein
Pac-pchgms_005	47	putative multidrug resistance protein ABC transporter family
Pac-pchgms_005	48	putative transferase family protein
Pac-pchgms_005	49	putative transferase/ transferase, transferring acyl groups other than amino-acyl groups
Pac-pchgms_006	01	putative polyprotein
Pac-pchgms_006	02	hypothetical protein
Pac-pchgms_006	03	putative submergence induced protein 2A
Pac-pchgms_006	04	hypothetical protein
Pac-pchgms_006	05	hypothetical protein
Pac-pchgms_006	06	putative polyprotein
Pac-pchgms_006	07	putative transducer HtrVI
Pac-pchgms_006	08	putative hAT family dimerisation domain containing protein
Pac-pchgms_006	09	putative phosphosugar-binding transcriptional regulator, RpiR family
Pac-pchgms_006	10	putative transferase/ transferase, transferring acyl groups other than amino-acyl groups
Pac-pchgms_006	11	putative transferase/ transferase, transferring acyl groups other than amino-acyl groups
Pac-pchgms_006	12	putative transferase/ transferase, transferring acyl groups other than amino-acyl groups
Pac-pchgms_006	13	hypothetical protein
Pac-pchgms_006	14	hypothetical protein
Pac-pchgms_006	15	putative synaptojanin 1, isoform CRA_e
Pac-pchgms_006	16	putative GAG-POL precursor
Pac-pchgms_006	17	putative GAG-POL precursor
Pac-pchgms_006	18	putative GAG-POL precursor
Pac-pchgms_006	19	putative GAG-POL precursor
Pac-pchgms_006	20	putative EEA1 (Early Endosome Antigen, Rab effector) homolog family member (eea-1)
Pac-pchgms_006	21	putative T-complex protein 1 subunit gamma
Pac-pchgms_006	22	putative TNP1
Pac-pchgms_006	23	putative TNP2
Pac-pchgms_006	24	putative TNP2
Pac-pchgms_006	25	putative aminoglycoside 3-N-acetyltransferase
Pac-pchgms_006	26	putative XRCC3; ATP binding / damaged DNA binding / protein binding / single-stranded DNA binding
Pac-pchgms_006	27	putative MuDRA-like transposase
Pac-pchgms_006	28	putative PRP38
Pac-pchgms_006	29	putative transposase
Pac-pchgms_006	30	putative UDP-glucose 4-epimerase
Pac-pchgms_006	31	putative ULP1A (UB-LIKE PROTEASE 1A); SUMO-specific protease/ cysteine-type peptidase
Pac-pchgms_006	32	putative transposase

Pac-pchgms_006	33	putative phage protein
Pac-pchgms_006	34	putative transposon protein Pong sub-class
Pac-pchgms_006	35	hypothetical protein similar to lipoprotein
Pac-pchgms_006	36	putative transferase/ transferase, transferring acyl groups other than amino-acyl groups
Pac-pchgms_006	37	hypothetical protein
Pac-pchgms_006	38	unknown protein
Pac-pchgms_006	39	putative transferase/ transferase, transferring acyl groups other than amino-acyl groups
Pac-pchgms_006	40	putative TNP1
Pac-pchgms_006	41	putative TNP2
Pac-pchgms_006	42	putative CXE carboxylesterase
Pac-pchgms_006	43	putative GRAS family transcription factor
Pac-pchgms_006	44	putative Metalloendopeptidase family-saccharolysin & thimet oligopeptidase (ISS)
Pac-pchgms_006	45	putative cytochrome P450 71 family protein
Pac-pchgms_006	46	putative chloroplast small heat shock protein
Pac-pchgms_006	47	putative ATPase
Pac-pchgms_006	48	putative Viral A-type inclusion protein repeat containing protein
Pac-pchgms_006	49	putative Mutator-like transposase
Pac-pchgms_006	50	hypothetical protein
Pac-pchgms_006	51	hypothetical protein
Pac-pchgms_006	52	putative transposase
Pac-pchgms_006	53	putative nubbin
Pac-pchgms_006	54	hypothetical protein
Pac-pchgms_006	55	putative polyprotein
Pac-pchgms_006	56	hypothetical protein
Pac-pchgms_006	57	unknown protein
Pac-pchgms_006	58	hypothetical protein
Pac-pchgms_006	59	hypothetical protein
Pac-pchgms_006	60	unknown protein
Pac-pchgms_006	61	hypothetical protein
Pac-pchgms_006	62	putative dopamine beta-monoxygenase
Pac-pchgms_007	01	putative A-kinase anchor protein 9 isoform 3
Pac-pchgms_007	02	putative Peptidoglycan-binding domain 1 protein
Pac-pchgms_007	03	hypothetical protein
Pac-pchgms_007	04	putative WDL1
Pac-pchgms_007	05	putative type-b response regulator
Pac-pchgms_007	06	putative E3 ubiquitin ligase
Pac-pchgms_007	07	putative leucine-rich repeat family protein / protein kinase family protein
Pac-pchgms_007	08	putative transposon protein
Pac-pchgms_007	09	putative TIR-NBS-LRR type disease resistance protein
Pac-pchgms_007	10	hypothetical protein
Pac-pchgms_007	11	putative protein binding / zinc ion binding
Pac-pchgms_007	12	putative CAAX amino terminal protease family protein
Pac-pchgms_007	13	putative NAD synthetase
Pac-pchgms_007	14	putative 3-ketoacyl-CoA reductase 2
Pac-pchgms_007	15	putative RNase H family protein
Pac-pchgms_007	16	putative Diapophytoene desaturase; AltName: Full=4,4'-diapophytoene desaturase
Pac-pchgms_007	17	putative CTV.20
Pac-pchgms_007	18	unknown protein
Pac-pchgms_007	19	putative 3-ketoacyl-CoA reductase 1
Pac-pchgms_007	20	unknown protein
Pac-pchgms_007	21	putative dentin sialophosphoprotein preproprotein
Pac-pchgms_007	22	putative methyltransferase

Pac-pchgms\_007 23 putative photosystem II core complex proteins psbY, chloroplast precursor  
 Pac-pchgms\_007 24 putative ALB4 (ALBINA 4)  
 Pac-pchgms\_007 25 hypothetical protein  
 Pac-pchgms\_007 26 putative regulator of chromosome condensation (RCC1) family protein  
 Pac-pchgms\_007 27 hypothetical protein  
 Pac-pchgms\_007 28 putative pectate lyase  
 Pac-pchgms\_007 29 putative T complex protein  
 Pac-pchgms\_007 30 hypothetical protein  
 Pac-pchgms\_007 31 putative RNA recognition motif (RRM)-containing protein  
 Pac-pchgms\_007 32 putative homeobox- domain containing protein  
 Pac-pchgms\_007 33 putative transposase  
 Pac-pchgms\_007 34 hypothetical protein  
 Pac-pchgms\_007 35 putative transposase  
 Pac-pchgms\_007 36 putative DIS1 (DISTORTED TRICHOMES 1); structural constituent of cytoskeleton  
 Pac-pchgms\_007 37 putative transposon protein Pong sub-class  
 Pac-pchgms\_007 38 hypothetical protein  
 Pac-pchgms\_007 39 putative oxysterol-binding protein  
 Pac-pchgms\_007 40 putative SDA1 family protein  
 Pac-pchgms\_007 41 putative catalytic/ pyridoxal phosphate binding  
 Pac-pchgms\_007 42 hypothetical protein  
 Pac-pchgms\_007 43 putative gag protein  
 Pac-pchgms\_007 44 hypothetical protein  
 Pac-pchgms\_007 45 putative terminal ear1-like 2 protein  
 Pac-pchgms\_007 46 unknown protein  
 Pac-pchgms\_007 47 putative cysteine proteinase  
 Pac-pchgms\_007 48 unknown protein  
 Pac-pchgms\_007 49 hypothetical protein  
 Pac-pchgms\_007 50 hypothetical protein  
 Pac-pchgms\_007 51 putative multidrug/pheromone exporter, MDR family, ABC transporter family  
 Pac-pchgms\_007 52 hypothetical protein  
 Pac-pchgms\_007 53 putative Integrase core domain containing protein  
 Pac-pchgms\_007 54 hypothetical protein  
 Pac-pchgms\_007 55 putative copia-like polyprotein  
 Pac-pchgms\_007 56 putative Ppx/GppA phosphatase family protein  
 Pac-pchgms\_007 57 putative multidrug/pheromone exporter, MDR family, ABC transporter family  
 Pac-pchgms\_007 58 putative multidrug/pheromone exporter, MDR family, ABC transporter family  
 Pac-pchgms\_008 01 hypothetical protein  
 Pac-pchgms\_008 02 putative polyprotein  
 Pac-pchgms\_008 03 putative integrase  
 Pac-pchgms\_008 04 putative transducin family protein / WD-40 repeat family protein  
 Pac-pchgms\_008 05 putative aminopeptidase N  
 Pac-pchgms\_008 06 putative sieve element-occluding protein 3  
 Pac-pchgms\_008 07 putative sieve element-occluding protein 3  
 Pac-pchgms\_008 08 putative forisome  
 Pac-pchgms\_008 09 putative pol-polyprotein  
 Pac-pchgms\_008 10 putative R27-2 protein  
 Pac-pchgms\_008 11 putative cytochrome P450  
 Pac-pchgms\_008 12 putative glycosyl hydrolase family 5 protein / cellulase family protein  
 Pac-pchgms\_008 13 putative polyadenylated-RNA export factor  
 Pac-pchgms\_008 14 putative RING-finger protein  
 Pac-pchgms\_008 15 hypothetical protein  
 Pac-pchgms\_008 16 putative viral A-type inclusion protein

Pac-pchgms\_008 17 putative cytochrome P450  
 Pac-pchgms\_008 18 putative site-specific DNA-methyltransferase (adenine-specific)  
 Pac-pchgms\_008 19 hypothetical protein  
 Pac-pchgms\_008 20 putative transposase  
 Pac-pchgms\_008 21 hypothetical protein  
 Pac-pchgms\_008 22 putative cytochrome P450  
 Pac-pchgms\_008 23 putative fatty oxidation complex alpha subunit  
 Pac-pchgms\_008 24 putative Mutator-like transposase  
 Pac-pchgms\_008 25 putative PK12 protein kinase  
 Pac-pchgms\_008 26 putative pentatricopeptide (PPR) repeat-containing protein  
 Pac-pchgms\_008 27 putative MAP kinase phosphatase 1  
 Pac-pchgms\_008 28 putative glutamic acid-rich protein cNBL1700  
 Pac-pchgms\_008 29 putative AlaT1  
 Pac-pchgms\_008 30 hypothetical protein similar to Tcc1j12.4  
 Pac-pchgms\_008 31 hypothetical protein  
 Pac-pchgms\_008 32 putative F-box family protein  
 Pac-pchgms\_008 33 putative polyphenol oxidase 2 precursor  
 Pac-pchgms\_008 34 hypothetical protein  
 Pac-pchgms\_008 35 putative copia-type polyprotein  
 Pac-pchgms\_008 36 hypothetical protein  
 Pac-pchgms\_008 37 putative polyphenol oxidase 2 precursor  
 Pac-pchgms\_008 38 hypothetical protein  
 Pac-pchgms\_008 39 hypothetical protein  
 Pac-pchgms\_008 40 hypothetical protein  
 Pac-pchgms\_008 41 hypothetical protein  
 Pac-pchgms\_008 42 hypothetical protein  
 Pac-pchgms\_008 43 putative ARO4 (ARMADILLO REPEAT ONLY 4); binding  
 Pac-pchgms\_008 44 putative peroxisomal copper-containing amine oxidase  
 Pac-pchgms\_008 45 putative ABI3-interacting protein 2  
 Pac-pchgms\_008 46 putative MATE efflux family protein  
 Pac-pchgms\_008 47 putative 60S ribosomal protein L27A  
 Pac-pchgms\_008 48 putative ATP-dependent protease Clp ATPase subunit  
 Pac-pchgms\_008 49 putative ATTAP1; ATPase, coupled to transmembrane movement of substances / transporter  
 Pac-pchgms\_008 50 putative PHD zinc finger-containing protein  
 Pac-pchgms\_008 51 hypothetical protein  
 Pac-pchgms\_008 52 hypothetical protein  
 Pac-pchgms\_008 53 putative FKBP-type peptidyl-prolyl cis-trans isomerases 1  
 Pac-pchgms\_008 54 putative glycosyl transferase  
 Pac-pchgms\_009 01 putative transposon protein Pong sub-class  
 Pac-pchgms\_009 02 putative DNA binding  
 Pac-pchgms\_009 03 hypothetical protein  
 Pac-pchgms\_009 04 hypothetical protein  
 Pac-pchgms\_009 05 putative octicosapeptide/Phox/Bem1p (PB1) domain-containing protein  
 Pac-pchgms\_009 06 hypothetical protein  
 Pac-pchgms\_009 07 putative transposon protein  
 Pac-pchgms\_009 08 hypothetical protein  
 Pac-pchgms\_009 09 putative thiol-disulfide isomerase-like thioredoxin  
 Pac-pchgms\_009 10 putative flavonol synthase  
 Pac-pchgms\_009 11 putative aldo/keto reductase  
 Pac-pchgms\_009 12 hypothetical protein  
 Pac-pchgms\_009 13 putative Mak16 protein  
 Pac-pchgms\_009 14 putative leucine-rich repeat family protein

Pac-pchgms\_009 15 hypothetical protein  
 Pac-pchgms\_009 16 hypothetical protein  
 Pac-pchgms\_009 17 hypothetical protein similar to GE21114  
 Pac-pchgms\_009 18 putative hAT family dimerisation domain containing protein  
 Pac-pchgms\_009 19 putative aldo/keto reductase  
 Pac-pchgms\_009 20 putative hAT family dimerisation domain containing protein  
 Pac-pchgms\_009 21 putative intracellular protein transport protein USO1  
 Pac-pchgms\_009 22 putative Mutator-like transposase  
 Pac-pchgms\_009 23 putative pol polyprotein  
 Pac-pchgms\_009 24 putative NADH dehydrogenase subunit J  
 Pac-pchgms\_009 25 putative retrotransposon protein  
 Pac-pchgms\_009 26 hypothetical protein  
 Pac-pchgms\_009 27 putative Mutator-like transposase  
 Pac-pchgms\_009 28 putative chemotaxis protein CheA  
 Pac-pchgms\_009 29 putative aldo/keto reductase  
 Pac-pchgms\_009 30 putative potential intra-Golgi transport complex subunit 7  
 Pac-pchgms\_009 31 hypothetical protein  
 Pac-pchgms\_009 32 putative AP2/ERF domain-containing transcription factor  
 Pac-pchgms\_009 33 putative CTV.20  
 Pac-pchgms\_009 34 unknown protein similar to heat shock protein DnaJ  
 Pac-pchgms\_009 35 putative phototropic-responsive NPH3 family protein  
 Pac-pchgms\_009 36 hypothetical protein  
 Pac-pchgms\_009 37 hypothetical protein  
 Pac-pchgms\_009 38 putative glycosyl transferase family 17 protein  
 Pac-pchgms\_009 39 putative homeobox protein C11b  
 Pac-pchgms\_009 40 putative TraG/TraD family protein  
 Pac-pchgms\_009 41 putative protein binding / zinc ion binding  
 Pac-pchgms\_009 42 putative glycosyltransferase involved in LPS biosynthesis  
 Pac-pchgms\_009 43 putative ATP-binding protein  
 Pac-pchgms\_009 44 putative protein phosphatase 2C  
 Pac-pchgms\_009 45 putative glutamyl-tRNA synthetase  
 Pac-pchgms\_009 46 putative GDSL-motif lipase/hydrolase family protein  
 Pac-pchgms\_009 47 putative GDSL-motif lipase/hydrolase family protein  
 Pac-pchgms\_009 48 putative SUFE2 (SULFUR E 2); enzyme activator  
 Pac-pchgms\_009 49 putative glycosyl hydrolase family protein 27 / alpha-galactosidase family protein / melibiase family protein  
 Pac-pchgms\_009 50 putative auxin-independent growth protein  
 Pac-pchgms\_009 51 putative RecName: Full=Ribose-phosphate pyrophosphokinase 3, mitochondrial  
 Pac-pchgms\_009 52 putative Glycoside hydrolase, family 28  
 Pac-pchgms\_009 53 putative polypeptide with a gag-like domain  
 Pac-pchgms\_009 54 putative pol protein  
 Pac-pchgms\_009 55 putative retrotransposon protein  
 Pac-pchgms\_010 01 putative Glycosyl transferase family, helical bundle domain protein  
 Pac-pchgms\_010 02 putative poor homologous synapsis 1 protein  
 Pac-pchgms\_010 03 putative peroxidase 1  
 Pac-pchgms\_010 04 putative ryanodine receptor RyR1 isoform  
 Pac-pchgms\_010 05 putative Chain A, Crystal Structure Of Highly Glycosylated Peroxidase From Royal Palm Tree  
 Pac-pchgms\_010 06 putative Protein kinase  
 Pac-pchgms\_010 07 putative protein kinase family protein / peptidoglycan-binding LysM domain-containing protein  
 Pac-pchgms\_010 08 putative polyprotein  
 Pac-pchgms\_010 09 putative DNA-directed RNA polymerase II largest subunit  
 Pac-pchgms\_010 10 putative HAP13 (HAPLESS 13); protein binding

Pac-pchgms\_010 11 putative u5 small nuclear ribonucleoprotein-specific protein  
 Pac-pchgms\_010 12 putative TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS 1  
 Pac-pchgms\_010 13 putativeL-tryptophan:2-oxoglutarate aminotransferase aminotransferase  
 Pac-pchgms\_010 14 putative FAR1; Zinc finger, SWIM-type  
 Pac-pchgms\_010 15 putative lipase  
 Pac-pchgms\_010 16 putative Zinc finger, CCHC-type  
 Pac-pchgms\_010 17 putative pentatricopeptide (PPR) repeat-containing protein  
 Pac-pchgms\_010 18 putative carboxyl-terminal proteinase  
 Pac-pchgms\_010 19 hypothetical protein  
 Pac-pchgms\_010 20 putative heme-binding protein 2  
 Pac-pchgms\_010 21 putative starch associated protein R1  
 Pac-pchgms\_010 22 hypothetical protein  
 Pac-pchgms\_010 23 hypothetical protein  
 Pac-pchgms\_010 24 putative Rehd high-affinity nitrate transporter NRT2.5  
 Pac-pchgms\_010 25 putative EDA24 (embryo sac development arrest 24); enzyme inhibitor pectinesterase inhibitor  
 Pac-pchgms\_010 26 putative polygalacturonase precursor homologue  
 Pac-pchgms\_010 27 putative protein kinase family protein  
 Pac-pchgms\_010 28 putative calcium channel, voltage-dependent, N type, alpha 1B subunit  
 Pac-pchgms\_010 29 hypothetical protein  
 Pac-pchgms\_010 30 putative protein kinase family protein  
 Pac-pchgms\_010 31 putative knotted1-like homeobox transcription factor  
 Pac-pchgms\_010 32 hypothetical protein  
 Pac-pchgms\_010 33 unknown protein  
 Pac-pchgms\_010 34 hypothetical protein  
 Pac-pchgms\_010 35 putative alpha galactosidase precursor  
 Pac-pchgms\_010 36 unknown protein  
 Pac-pchgms\_010 37 putative leucine-rich repeat receptor-like kinase  
 Pac-pchgms\_010 38 putative F-box/kelch protein  
 Pac-pchgms\_010 39 putative ATP citrate lyase b-subunit  
 Pac-pchgms\_010 40 putative protein binding / zinc ion binding  
 Pac-pchgms\_010 41 putative RecName: Full=DNA-directed RNA polymerases I, II, and III subunit RPABC5;  
 Pac-pchgms\_010 42 hypothetical protein  
 Pac-pchgms\_010 43 putative INO (INNER NO OUTER); protein binding / transcription factor  
 Pac-pchgms\_010 44 putative ubiquitin extension protein  
 Pac-pchgms\_010 45 putative CAF2; RNA splicing factor, transesterification mechanism  
 Pac-pchgms\_010 46 putative HupE / UreJ protein  
 Pac-pchgms\_010 47 putative RecName: Full=Cytochrome b-c1 complex subunit 8;  
 Pac-pchgms\_010 48 hypothetical protein  
 Pac-pchgms\_010 49 putative polygalacturonase  
 Pac-pchgms\_011 01 hypothetical protein  
 Pac-pchgms\_011 02 putative polygalacturonase  
 Pac-pchgms\_011 03 putative ECHID (ENOYL-COA HYDRATASE/ISOMERASE D); catalytic/ naphthoate synthase  
 Pac-pchgms\_011 04 putative RecName: Full=Phosphatidylinositol 3-kinase, root isoform; SPI3K-5  
 Pac-pchgms\_011 05 putative ADP-ribosylation factor  
 Pac-pchgms\_011 06 putative galactinol synthase 1  
 Pac-pchgms\_011 07 putative Acetyltransferase, GNAT family  
 Pac-pchgms\_011 08 putative ATPANK2 (PANTOTHENATE KINASE 2); pantothenate kinase  
 Pac-pchgms\_011 09 putative OBP32pep  
 Pac-pchgms\_011 10 putative OBP32pep  
 Pac-pchgms\_011 11 putative Viral A-type inclusion protein repeat containing protein  
 Pac-pchgms\_011 12 putative Mutator-like transposase  
 Pac-pchgms\_011 13 putative protein disulfide isomerase (PDI)-like protein 3

Pac-pchgms\_011 14 putative DC1 domain-containing protein  
 Pac-pchgms\_011 15 hypothetical protein  
 Pac-pchgms\_011 16 hypothetical protein  
 Pac-pchgms\_011 17 putative DC1 domain-containing protein  
 Pac-pchgms\_011 18 hypothetical protein similar to predicted transcriptional regulator with HTH domain  
 Pac-pchgms\_011 19 putative beta-1,3-glucanase  
 Pac-pchgms\_011 20 putative acetylglutamate kinase  
 Pac-pchgms\_011 21 hypothetical protein  
 Pac-pchgms\_011 22 putative DNA binding / transcription factor  
 Pac-pchgms\_011 23 putative protein kinase family protein  
 Pac-pchgms\_011 24 putative reverse transcriptase  
 Pac-pchgms\_011 25 hypothetical protein  
 Pac-pchgms\_011 26 putative Poly(ADP-ribose) polymerase, catalytic region  
 Pac-pchgms\_011 27 putative protein kinase  
 Pac-pchgms\_011 28 putative synaptojanin 2 isoform a  
 Pac-pchgms\_011 29 putative calmodulin-binding protein  
 Pac-pchgms\_011 30 putative mov34/MPN/PAD-1 family protein  
 Pac-pchgms\_011 31 putative TNP2  
 Pac-pchgms\_011 32 putative TNP1  
 Pac-pchgms\_011 33 putative chloroplast carbonic anhydrase  
 Pac-pchgms\_011 34 putative Ulp1-like peptidase  
 Pac-pchgms\_011 35 putative F-box family protein  
 Pac-pchgms\_011 36 putative flagellar inner arm dynein light chain p28  
 Pac-pchgms\_011 37 putative HsdR family type I site-specific deoxyribonuclease  
 Pac-pchgms\_011 38 putative phage major tail tube protein  
 Pac-pchgms\_011 39 putative oxidoreductase, zinc-binding dehydrogenase family protein  
 Pac-pchgms\_011 40 putative OB-fold nucleic acid binding domain containing protein  
 Pac-pchgms\_011 41 hypothetical protein  
 Pac-pchgms\_011 42 putative polygalacturonase isoenzyme 1 beta subunit homolog  
 Pac-pchgms\_011 43 putative peptidase M20  
 Pac-pchgms\_011 44 hypothetical protein  
 Pac-pchgms\_011 45 putative polygalacturonase isoenzyme 1 beta subunit homolog  
 Pac-pchgms\_011 46 hypothetical protein  
 Pac-pchgms\_011 47 putative peptidase M29, aminopeptidase II  
 Pac-pchgms\_011 48 putative Legume lectin, beta domain  
 Pac-pchgms\_011 49 putative transducin family protein / WD-40 repeat family protein  
 Pac-pchgms\_011 50 putative metal-dependent phosphohydrolase HD domain-containing protein  
 Pac-pchgms\_011 51 putative protein kinase  
 Pac-pchgms\_011 52 putative MPF1-like-B  
 Pac-pchgms\_011 53 putative TNP2  
 Pac-pchgms\_011 54 putative tnp1 protein  
 Pac-pchgms\_011 55 putative truncated copia-type polyprotein  
 Pac-pchgms\_012 01 hypothetical protein  
 Pac-pchgms\_012 02 putative TNP2  
 Pac-pchgms\_012 03 putative tnp1 protein  
 Pac-pchgms\_012 04 hypothetical protein similar to GL20871  
 Pac-pchgms\_012 05 putative ATP binding protein  
 Pac-pchgms\_012 06 putative ULP1D (UB-LIKE PROTEASE 1D); SUMO-specific protease/ cysteine-type peptidase  
 Pac-pchgms\_012 07 putative N-acetylmuramoyl-L-alanine amidase domain-containing protein  
 Pac-pchgms\_012 08 hypothetical protein  
 Pac-pchgms\_012 09 putative splicing factor PWI domain-containing protein / RNA recognition motif (RRM)-containing protein

Pac-pchgms_012	10	putative ENT1,AT (EQUILIBRATIVE NUCLEOTIDE TRANSPORTER 1); nucleoside transmembrane transporter
Pac-pchgms_012	11	putative two-component system sensor histidine kinase/response regulator hybrid
Pac-pchgms_012	12	putative serine/threonine kinase protein
Pac-pchgms_012	13	putative armadillo/beta-catenin repeat family protein / U-box domain-containing protein
Pac-pchgms_012	14	hypothetical protein
Pac-pchgms_012	15	putative xyloglucan endotransglycosylase/hydrolase precursor XTH-38
Pac-pchgms_012	16	putative aldehyde dehydrogenase (NAD+)
Pac-pchgms_012	17	unknown protein
Pac-pchgms_012	18	hypothetical protein similar to GA19937
Pac-pchgms_012	19	putative spermidine synthase
Pac-pchgms_012	20	putative ABC transporter, membrane spanning protein (sugar/ribonucleotide)
Pac-pchgms_012	21	putative carotenoid cleavage dioxygenase 4
Pac-pchgms_012	22	putative RSZP21 (RS-CONTAINING ZINC FINGER PROTEIN 21); protein binding
Pac-pchgms_012	23	putative potassium transporter family protein
Pac-pchgms_012	24	putative potassium transporter
Pac-pchgms_012	25	hypothetical protein
Pac-pchgms_012	26	putative trehalose 6-phosphate synthase
Pac-pchgms_012	27	hypothetical protein
Pac-pchgms_012	28	putative NHL repeat-containing protein
Pac-pchgms_012	29	putative NHL repeat-containing protein
Pac-pchgms_012	30	putative binding / clathrin binding / protein binding / protein transporter
Pac-pchgms_012	31	putative RING finger protein
Pac-pchgms_012	32	putative threonyl-tRNA synthetase
Pac-pchgms_012	33	putative transcription factor
Pac-pchgms_012	34	hypothetical protein
Pac-pchgms_012	35	putative Ulp1-like peptidase
Pac-pchgms_012	36	hypothetical protein
Pac-pchgms_012	37	putative ArsR family transcriptional regulator
Pac-pchgms_012	38	putative FAR1; Zinc finger, SWIM-type
Pac-pchgms_012	39	putative nodulin MtN21 family protein
Pac-pchgms_012	40	hypothetical protein
Pac-pchgms_012	41	putative nodulin MtN21 family protein
Pac-pchgms_012	42	putative nodulin MtN21 family protein
Pac-pchgms_012	43	putative nodulin MtN21 family protein
Pac-pchgms_012	44	hypothetical protein
Pac-pchgms_012	45	hypothetical protein
Pac-pchgms_012	46	hypothetical protein
Pac-pchgms_012	47	putative nodulin MtN21 family protein
Pac-pchgms_012	48	hypothetical protein
Pac-pchgms_012	49	putative nodulin MtN21 family protein
Pac-pchgms_012	50	putative nodulin MtN21 family protein
Pac-pchgms_013	01	hypothetical protein
Pac-pchgms_013	02	putative transposase
Pac-pchgms_013	03	hypothetical protein
Pac-pchgms_013	04	putative MLP423 (MLP-LIKE PROTEIN 423)
Pac-pchgms_013	05	hypothetical protein
Pac-pchgms_013	06	putative GagPol3
Pac-pchgms_013	07	putative GAG-POL precursor
Pac-pchgms_013	08	hypothetical protein
Pac-pchgms_013	09	putative protein kinase family protein
Pac-pchgms_013	10	putative YLS7

Pac-pchgms\_013 11 putative AGR340Wp  
 Pac-pchgms\_013 12 putative TIR-NBS-LRR type disease resistance protein  
 Pac-pchgms\_013 13 putative nucleobase ascorbate transporter  
 Pac-pchgms\_013 14 hypothetical protein  
 Pac-pchgms\_013 15 putative Protein prenyltransferase alpha subunit repeat containing protein  
 Pac-pchgms\_013 16 putative tetrapyrrole (corrin/porphyrin) methylase  
 Pac-pchgms\_013 17 hypothetical protein  
 Pac-pchgms\_013 18 putative N-acetyl-glutamate synthase  
 Pac-pchgms\_013 19 putative cyclin D1  
 Pac-pchgms\_013 20 putative centromere protein F  
 Pac-pchgms\_013 21 putative RNase H family protein  
 Pac-pchgms\_013 22 putative retrotransposon protein  
 Pac-pchgms\_013 23 putative transposon protein Pong sub-class  
 Pac-pchgms\_013 24 putative RNase H domain-containing protein  
 Pac-pchgms\_013 25 hypothetical protein  
 Pac-pchgms\_013 26 putative DNA polymerase lambda (POLL)  
 Pac-pchgms\_013 27 putative RecName: Full=DEAD-box ATP-dependent RNA helicase 22  
 Pac-pchgms\_013 28 putative UDP-glucose:glucosyltransferase  
 Pac-pchgms\_013 29 hypothetical protein  
 Pac-pchgms\_013 30 putative Progesterone-induced-blocking factor  
 Pac-pchgms\_013 31 hypothetical protein  
 Pac-pchgms\_013 32 putative chloroplast envelope protein 1  
 Pac-pchgms\_013 33 putative iaa-amino acid hydrolase 10  
 Pac-pchgms\_013 34 putative hAT family dimerisation domain containing protein, expressed  
 Pac-pchgms\_013 35 putative peroxidase  
 Pac-pchgms\_013 36 putative ARL1 (ARG1-LIKE 1); heat shock protein binding / unfolded protein binding  
 Pac-pchgms\_013 37 hypothetical protein  
 Pac-pchgms\_013 38 hypothetical protein  
 Pac-pchgms\_013 39 putative zinc finger protein 5, ZFP5  
 Pac-pchgms\_013 40 hypothetical protein  
 Pac-pchgms\_013 41 putative chloroplast envelope protein 1  
 Pac-pchgms\_013 42 putative SAM domain family protein  
 Pac-pchgms\_013 43 hypothetical protein  
 Pac-pchgms\_013 44 hypothetical protein similar to Os04g0517400  
 Pac-pchgms\_013 45 hypothetical protein  
 Pac-pchgms\_013 46 putative matrix metalloprotease 1  
 Pac-pchgms\_013 47 putative UVI1  
 Pac-pchgms\_013 48 hypothetical protein  
 Pac-pchgms\_013 49 putative UL13  
 Pac-pchgms\_013 50 putative zinc ion binding  
 Pac-pchgms\_013 51 putative ARR3 (RESPONSE REGULATOR 3); transcription regulator/ two-component response regulator  
 Pac-pchgms\_014 01 putative hairpin-inducing protein  
 Pac-pchgms\_014 02 putative LL20 15kDa ladder antigen  
 Pac-pchgms\_014 03 hypothetical protein similar to Os08g0425000  
 Pac-pchgms\_014 04 putative Respiratory nitrate reductase gamma chain  
 Pac-pchgms\_014 05 hypothetical protein  
 Pac-pchgms\_014 06 putative TNP1  
 Pac-pchgms\_014 07 putative heat shock protein Hsp20 domain-containing protein  
 Pac-pchgms\_014 08 putative ATFH8 (formin 8); actin binding / actin filament binding / profilin binding  
 Pac-pchgms\_014 09 putative WDL1  
 Pac-pchgms\_014 10 hypothetical protein

Pac-pchgms\_014 11 putative germin-like protein 5  
 Pac-pchgms\_014 12 putative glycosyltransferase, CAZy family GT8  
 Pac-pchgms\_014 13 hypothetical protein  
 Pac-pchgms\_014 14 putative pyruvate dehydrogenase alpha subunit  
 Pac-pchgms\_014 15 putative EMB25 (EMBRYO DEFECTIVE 25); ATP-dependent helicase/ RNA helicase  
 Pac-pchgms\_014 16 putative SIN3 component, histone deacetylase complex  
 Pac-pchgms\_014 17 unknown protein  
 Pac-pchgms\_014 18 hypothetical protein  
 Pac-pchgms\_014 19 putative binding  
 Pac-pchgms\_014 20 putative Mutator-like transposase  
 Pac-pchgms\_014 21 putative zinc finger (C3HC4-type RING finger) family protein  
 Pac-pchgms\_014 22 putative CCB4 (COFACTOR ASSEMBLY OF COMPLEX C)  
 Pac-pchgms\_014 23 hypothetical protein  
 Pac-pchgms\_014 24 putative MYB transcription factor MYB138  
 Pac-pchgms\_014 25 putative ATP-dependent RNA helicase RhIE  
 Pac-pchgms\_014 26 putative receptor-like protein kinase  
 Pac-pchgms\_014 27 putative UBA/THIF-type NAD/FAD binding protein  
 Pac-pchgms\_014 28 putative type 2A protein phosphatase-1  
 Pac-pchgms\_014 29 putative peptidase, M23/M37 family protein  
 Pac-pchgms\_014 30 putative DNA binding  
 Pac-pchgms\_014 31 putative regulatory protein  
 Pac-pchgms\_014 32 putative TSPY-like 5  
 Pac-pchgms\_014 33 putative short chain dehydrogenase  
 Pac-pchgms\_014 34 putative SIP1b  
 Pac-pchgms\_014 35 putative 5-formyltetrahydrofolate cyclo-ligase  
 Pac-pchgms\_014 36 hypothetical protein  
 Pac-pchgms\_014 37 putative transposase  
 Pac-pchgms\_014 38 putative dihydrolipoyllysine-residue succinyltransferase, E2 component  
 Pac-pchgms\_014 39 putative tir-nbs-lrr resistance protein  
 Pac-pchgms\_014 40 putative tir-nbs-lrr resistance protein  
 Pac-pchgms\_014 41 hypothetical protein  
 Pac-pchgms\_014 42 putative RNA-directed DNA polymerase (Reverse transcriptase)  
 Pac-pchgms\_014 43 putative Mutator-like transposase  
 Pac-pchgms\_014 44 putative tir-nbs-lrr resistance protein  
 Pac-pchgms\_014 45 putative aminophospholipid ATPase  
 Pac-pchgms\_014 46 putative aminophospholipid ATPase  
 Pac-pchgms\_014 47 putative ZCW32; DNA binding / transcription factor  
 Pac-pchgms\_014 48 putative viral A-type inclusion protein  
 Pac-pchgms\_014 49 putative signal transduction histidine kinase, LytS  
 Pac-pchgms\_014 50 putative CW<sup>14</sup>  
 Pac-pchgms\_014 51 putative disease resistance family protein / LRR family protein  
 Pac-pchgms\_014 52 putative ER glycerol-phosphate acyltransferase  
 Pac-pchgms\_015 01 putative galactosyltransferase family protein  
 Pac-pchgms\_015 02 putative transposase  
 Pac-pchgms\_015 03 hypothetical protein  
 Pac-pchgms\_015 04 hypothetical protein  
 Pac-pchgms\_015 05 putative P30Sh95F04  
 Pac-pchgms\_015 06 putative gag protein  
 Pac-pchgms\_015 07 hypothetical protein  
 Pac-pchgms\_015 08 putative potential GRIP domain Golgi protein  
 Pac-pchgms\_015 09 putative gag-pol polyprotein  
 Pac-pchgms\_015 10 putative tau class glutathione transferase GSTU51

Pac-pchgms_015	11	putative glutathione S-transferase 3
Pac-pchgms_015	12	putative tau class glutathione transferase GSTU51
Pac-pchgms_015	13	putative binding
Pac-pchgms_015	14	putative protein kinase family protein
Pac-pchgms_015	15	putative ThiF family protein
Pac-pchgms_015	16	putative CRR28 (CHLORORESPIRATORY REDUCTION28); endonuclease
Pac-pchgms_015	17	putative Bbc1p
Pac-pchgms_015	18	putative transposase
Pac-pchgms_015	19	putative potyviral capsid protein interacting protein 2a
Pac-pchgms_015	20	putative thioredoxin h
Pac-pchgms_015	21	hypothetical protein
Pac-pchgms_015	22	putative alanine racemase
Pac-pchgms_015	23	putative FAC1 (EMBRYONIC FACTOR1); AMP deaminase
Pac-pchgms_015	24	putative pol polyprotein
Pac-pchgms_015	25	putative TNP2
Pac-pchgms_015	26	hypothetical protein
Pac-pchgms_015	27	hypothetical protein
Pac-pchgms_015	28	hypothetical protein
Pac-pchgms_015	29	putative glycosyltransferase
Pac-pchgms_015	30	hypothetical protein
Pac-pchgms_015	31	putative proton-dependent oligopeptide transport (POT) family protein
Pac-pchgms_015	32	putative proton-dependent oligopeptide transport (POT) family protein
Pac-pchgms_015	33	hypothetical protein
Pac-pchgms_015	34	putative GAG-POL precursor
Pac-pchgms_015	35	hypothetical protein
Pac-pchgms_015	36	putative proton-dependent oligopeptide transport (POT) family protein
Pac-pchgms_015	37	hypothetical protein
Pac-pchgms_015	38	putative RNA-directed DNA polymerase homolog T13L16.7
Pac-pchgms_015	39	putative proton-dependent oligopeptide transport (POT) family protein
Pac-pchgms_015	40	hypothetical protein
Pac-pchgms_015	41	putative Isoamylase N-terminal domain containing protein, expressed
Pac-pchgms_015	42	putative regulator of chromosome condensation (RCC1) family protein
Pac-pchgms_015	43	putative small multi-drug export protein
Pac-pchgms_015	44	putative homeodomain leucine zipper protein HDZ2
Pac-pchgms_015	45	putative pentatricopeptide (PPR) repeat-containing protein
Pac-pchgms_015	46	putative NADH dehydrogenase subunit 4L
Pac-pchgms_015	47	putative nitrate transporter (NTL1); 53025-56402
Pac-pchgms_015	48	putative glyoxal oxidase
Pac-pchgms_015	49	putative Integrase core domain containing protein
Pac-pchgms_015	50	putative PLPB (PAS/LOV PROTEIN B); signal transducer/ two-component sensor
Pac-pchgms_015	51	putative glycosyltransferase 36
Pac-pchgms_015	52	hypothetical protein
Pac-pchgms_015	53	putative serine/threonine protein kinase
Pac-pchgms_015	54	putative bifunctional N-succinyldiaminopimelate- aminotransferase/acetylornithine transaminase protein
Pac-pchgms_016	01	hypothetical protein
Pac-pchgms_016	02	unknown protein
Pac-pchgms_016	03	hypothetical protein
Pac-pchgms_016	04	putative hypersensitive-induced response protein
Pac-pchgms_016	05	putative membrane alanine aminopeptidase
Pac-pchgms_016	06	hypothetical protein
Pac-pchgms_016	07	hypothetical protein

Pac-pchgms\_016 08 putative porin  
 Pac-pchgms\_016 09 putative plastid alpha-amylase  
 Pac-pchgms\_016 10 hypothetical protein  
 Pac-pchgms\_016 11 putative DNA binding  
 Pac-pchgms\_016 12 putative transposase  
 Pac-pchgms\_016 13 putative RNA-directed DNA polymerase homolog T13L16.7  
 Pac-pchgms\_016 14 putative CheA signal transduction histidine kinase  
 Pac-pchgms\_016 15 hypothetical protein  
 Pac-pchgms\_016 16 putative pectate lyase homolog  
 Pac-pchgms\_016 17 putative chromosome segregation protein SMC  
 Pac-pchgms\_016 18 putative hydroxyproline-rich glycoprotein family protein  
 Pac-pchgms\_016 19 putative TIR-NBS-LRR type disease resistance protein  
 Pac-pchgms\_016 20 hypothetical protein  
 Pac-pchgms\_016 21 hypothetical protein  
 Pac-pchgms\_016 22 unknown protein  
 Pac-pchgms\_016 23 putative AT59; lyase/ pectate lyase  
 Pac-pchgms\_016 24 putative AP2 domain-containing transcription factor  
 Pac-pchgms\_016 25 putative bZIP transcription factor bZIP109  
 Pac-pchgms\_016 26 putative zinc finger protein  
 Pac-pchgms\_016 27 hypothetical protein  
 Pac-pchgms\_016 28 putative TNP1  
 Pac-pchgms\_016 29 putative TNP2  
 Pac-pchgms\_016 30 putative TPA: TPA\_inf: WRKY transcription factor 73  
 Pac-pchgms\_016 31 putative polyprotein  
 Pac-pchgms\_016 32 putative transposon protein Pong sub-class  
 Pac-pchgms\_016 33 putative CBS domain-containing protein  
 Pac-pchgms\_016 34 putative CBS domain-containing protein  
 Pac-pchgms\_016 35 putative pentatricopeptide repeat-containing protein  
 Pac-pchgms\_016 36 hypothetical protein similar to retrotransposon protein  
 Pac-pchgms\_016 37 hypothetical protein  
 Pac-pchgms\_016 38 hypothetical protein  
 Pac-pchgms\_016 39 putative Terpenoid cyclases/protein prenyltransferase alpha-alpha toroid; Bacterial adhesion  
 Pac-pchgms\_016 40 putative histone ubiquitination proteins group  
 Pac-pchgms\_016 41 putative Protein kinase  
 Pac-pchgms\_016 42 putative CTV.20  
 Pac-pchgms\_016 43 putative CTV.20  
 Pac-pchgms\_016 44 putative IQD29 (IQ-domain 29); calmodulin binding  
 Pac-pchgms\_016 45 putative Pyrrolo-quinoline quinone  
 Pac-pchgms\_016 46 hypothetical protein  
 Pac-pchgms\_016 47 hypothetical protein  
 Pac-pchgms\_016 48 putative Gag-protease-integrase-RT-RNaseH polyprotein  
 Pac-pchgms\_016 49 putative copia-type polyprotein  
 Pac-pchgms\_016 50 putative CCHC-type integrase  
 Pac-pchgms\_016 51 putative integrase  
 Pac-pchgms\_016 52 putative truncated copia-type polyprotein  
 Pac-pchgms\_016 53 putative TWiK family of potassium channels family member (twk-2)  
 Pac-pchgms\_016 54 unknown protein  
 Pac-pchgms\_016 55 hypothetical protein similar to WD-repeat protein 12 (ISS)  
 Pac-pchgms\_016 56 hypothetical protein  
 Pac-pchgms\_016 57 putative zinc finger (C3HC4-type RING finger) family protein  
 Pac-pchgms\_016 58 hypothetical protein  
 Pac-pchgms\_016 59 putative carboxylate-amine ligase

Pac-pchgms\_016 60 putative zinc finger (C3HC4-type RING finger) family protein  
 Pac-pchgms\_016 61 hypothetical protein  
 Pac-pchgms\_017 01 hypothetical protein  
 Pac-pchgms\_017 02 hypothetical protein  
 Pac-pchgms\_017 03 unknown protein  
 Pac-pchgms\_017 04 putative APK2B (PROTEIN KINASE 2B); ATP binding / kinase/ protein kinase/ protein serine/threonine kinase  
 Pac-pchgms\_017 05 putative Protein kinase  
 Pac-pchgms\_017 06 putative UTR3 (UDP-GALACTOSE TRANSPORTER 3); pyrimidine nucleotide sugar transmembrane transporter  
 Pac-pchgms\_017 07 putative photosystem I assembly protein Ycf3  
 Pac-pchgms\_017 08 hypothetical protein  
 Pac-pchgms\_017 09 putative transposase  
 Pac-pchgms\_017 10 putative MYB88 (myb domain protein 88); DNA binding / transcription factor  
 Pac-pchgms\_017 11 hypothetical protein  
 Pac-pchgms\_017 12 hypothetical protein  
 Pac-pchgms\_017 13 putative oxidoreductase  
 Pac-pchgms\_017 14 putative HAHB-1  
 Pac-pchgms\_017 15 hypothetical protein  
 Pac-pchgms\_017 16 putative sucrose transporter 2B  
 Pac-pchgms\_017 17 putative RNA recognition motif (RRM)-containing protein  
 Pac-pchgms\_017 18 putative peptidyl-prolyl cis-trans isomerase cyclophilin-type family protein  
 Pac-pchgms\_017 19 hypothetical protein  
 Pac-pchgms\_017 20 putative CrcB-like family protein  
 Pac-pchgms\_017 21 hypothetical protein  
 Pac-pchgms\_017 22 hypothetical protein  
 Pac-pchgms\_017 23 unknown protein  
 Pac-pchgms\_017 24 putative kelch repeat-containing F-box family protein  
 Pac-pchgms\_017 25 putative TO109-12  
 Pac-pchgms\_017 26 hypothetical protein  
 Pac-pchgms\_017 27 hypothetical protein  
 Pac-pchgms\_017 28 putative pentatricopeptide (PPR) repeat-containing protein  
 Pac-pchgms\_017 29 putative Ran GTPase binding / chromatin binding / zinc ion binding  
 Pac-pchgms\_017 30 putative ATHVA22C  
 Pac-pchgms\_017 31 putative BHLH  
 Pac-pchgms\_017 32 putative Mog1 protein  
 Pac-pchgms\_017 33 putative ATCUL3 (ARABIDOPSIS THALIANA CULLIN 3); protein binding / ubiquitin-protein ligase  
 Pac-pchgms\_017 34 putative haloacid dehalogenase-like hydrolase family protein  
 Pac-pchgms\_017 35 putative pyruvate oxidase  
 Pac-pchgms\_017 36 hypothetical protein  
 Pac-pchgms\_017 37 putative MuDRA-like transposase  
 Pac-pchgms\_017 38 putative MuDRA-like transposase  
 Pac-pchgms\_017 39 putative Asparagine-rich protein  
 Pac-pchgms\_017 40 putative aminopeptidase P  
 Pac-pchgms\_017 41 hypothetical protein  
 Pac-pchgms\_017 42 putative Ulp1-like peptidase  
 Pac-pchgms\_017 43 putative binding  
 Pac-pchgms\_017 44 putative ORC6 (ORIGIN RECOGNITION COMPLEX PROTEIN 6); DNA binding  
 Pac-pchgms\_017 45 putative gag-pol polyprotein  
 Pac-pchgms\_017 46 putative SBH1 (SPHINGOID BASE HYDROXYLASE 1); catalytic/ sphingosine hydroxylase  
 Pac-pchgms\_017 47 putative dehydration-responsive family protein  
 Pac-pchgms\_017 48 hypothetical protein  
 Pac-pchgms\_017 49 putative CAAX amino terminal protease family protein

Pac-pchgms\_017 50 putative Cyclin-like F-box  
 Pac-pchgms\_017 51 putative copia-type polyprotein  
 Pac-pchgms\_017 52 hypothetical protein  
 Pac-pchgms\_017 53 hypothetical protein  
 Pac-pchgms\_017 54 putative Thioredoxin domain protein  
 Pac-pchgms\_017 55 putative nucleic acid binding / zinc ion binding  
 Pac-pchgms\_017 56 hypothetical protein similar to EMB2756 (EMBRYO DEFECTIVE 2756)  
 Pac-pchgms\_018 01 putative CAAX amino terminal protease family protein  
 Pac-pchgms\_018 02 putative NAC domain protein, IPR003441  
 Pac-pchgms\_018 03 putative glutathione S-transferase 12  
 Pac-pchgms\_018 04 putative 60S ribosomal protein L34 (RPL34A)  
 Pac-pchgms\_018 05 putative structural constituent of ribosome  
 Pac-pchgms\_018 06 putative zinc finger-homeodomain protein 1  
 Pac-pchgms\_018 07 putative CLE family OsCLE306 protein  
 Pac-pchgms\_018 08 putative zinc finger (C3HC4-type RING finger) family protein  
 Pac-pchgms\_018 09 putative nucleoside phosphatase family protein / GDA1/CD39 family protein  
 Pac-pchgms\_018 10 putative DNA-binding protein  
 Pac-pchgms\_018 11 putative amino acid transporter  
 Pac-pchgms\_018 12 putative phage head morphogenesis protein, SPP1 gp7 family  
 Pac-pchgms\_018 13 putative pentatricopeptide (PPR) repeat-containing protein  
 Pac-pchgms\_018 14 putative amino acid transporter  
 Pac-pchgms\_018 15 putative S-like ribonuclease  
 Pac-pchgms\_018 16 putative S-like ribonuclease  
 Pac-pchgms\_018 17 hypothetical protein  
 Pac-pchgms\_018 18 unknown protein  
 Pac-pchgms\_018 19 putative GALT1 (GALACTOSYLTRANSFERASE1); UDP-galactose:N-glycan beta-1,3-galactosyltransferase  
 Pac-pchgms\_018 20 putative zinc finger (C3HC4-type RING finger) family protein  
 Pac-pchgms\_018 21 putative glucose-methanol-choline (GMC) oxidoreductase family protein  
 Pac-pchgms\_018 22 putative transcription factor  
 Pac-pchgms\_018 23 putative gag-pol polymerase  
 Pac-pchgms\_018 24 putative pol-polyprotein  
 Pac-pchgms\_018 25 hypothetical protein  
 Pac-pchgms\_018 26 putative f-box family protein  
 Pac-pchgms\_018 27 putative calcium-binding EF hand family protein  
 Pac-pchgms\_018 28 hypothetical protein  
 Pac-pchgms\_018 29 putative myb family transcription factor (MYB117)  
 Pac-pchgms\_018 30 unknown protein  
 Pac-pchgms\_018 31 hypothetical protein  
 Pac-pchgms\_018 32 putative ORF III polyprotein  
 Pac-pchgms\_018 33 putative Mutator-like transposase  
 Pac-pchgms\_018 34 putative conserved Plasmodium protein  
 Pac-pchgms\_018 35 unknown protein  
 Pac-pchgms\_018 36 putative mitochondrial transcription termination factor  
 Pac-pchgms\_018 37 putative DNA binding / transcription factor  
 Pac-pchgms\_018 38 putative expansin  
 Pac-pchgms\_018 39 hypothetical protein  
 Pac-pchgms\_018 40 putative UbiE/COQ5 methyltransferase family protein  
 Pac-pchgms\_018 41 putative RecName: Full=UPF0497 membrane protein Os06g0231050  
 Pac-pchgms\_018 42 putative oxygen evolving enhancer 3 (PsbQ) family protein  
 Pac-pchgms\_018 43 putative DUF962 domain-containing protein  
 Pac-pchgms\_018 44 putative electron carrier/ heme binding / iron ion binding / monooxygenase/ oxygen binding

Pac-pchgms\_018 45 hypothetical protein  
 Pac-pchgms\_018 46 putative exocyst complex subunit Sec15-like family protein  
 Pac-pchgms\_018 47 putative integral membrane protein  
 Pac-pchgms\_018 48 putative integrase core domain containing protein  
 Pac-pchgms\_018 49 putative protein dimerization  
 Pac-pchgms\_018 50 hypothetical protein  
 Pac-pchgms\_018 51 putative transposase  
 Pac-pchgms\_018 52 putative emb1579 (embryo defective 1579); binding / calcium ion binding  
 Pac-pchgms\_018 53 putative SET domain protein  
 Pac-pchgms\_018 54 putative Mitochondrial carrier protein  
 Pac-pchgms\_018 55 putative MYB transcription factor ML2  
 Pac-pchgms\_018 56 putative Glycosyl hydrolases family 32  
 Pac-pchgms\_018 57 putative peptidyl-prolyl cis-trans isomerase PPIC-type family protein  
 Pac-pchgms\_018 58 putative alpha-1,2-fucosyltransferase  
 Pac-pchgms\_018 59 putative EXS family protein / ERD1/XPR1/SYG1 family protein  
 Pac-pchgms\_018 60 putative EXS family protein / ERD1/XPR1/SYG1 family protein  
 Pac-pchgms\_019 01 putative EXS family protein / ERD1/XPR1/SYG1 family protein  
 Pac-pchgms\_019 02 putative polyprotein  
 Pac-pchgms\_019 03 putative EXS family protein / ERD1/XPR1/SYG1 family protein  
 Pac-pchgms\_019 04 putative TNP2  
 Pac-pchgms\_019 05 putative EXS family protein / ERD1/XPR1/SYG1 family protein  
 Pac-pchgms\_019 06 putative Gag-Pol polyprotein  
 Pac-pchgms\_019 07 hypothetical protein  
 Pac-pchgms\_019 08 putative RecName: Full=Ribulose-1,5 bisphosphate carboxylase/oxygenase large subunit  
 Pac-pchgms\_019 09 putative auxin-independent growth promoter  
 Pac-pchgms\_019 10 putative emp24/gp25L/p24 family protein  
 Pac-pchgms\_019 11 putative VIK (VH1-INTERACTING KINASE);  
 Pac-pchgms\_019 12 putative polyprotein  
 Pac-pchgms\_019 13 putative polyprotein  
 Pac-pchgms\_019 14 putative 26S proteasome non-ATPase regulatory subunit 11  
 Pac-pchgms\_019 15 hypothetical protein  
 Pac-pchgms\_019 16 putative transposon protein Pong sub-class  
 Pac-pchgms\_019 17 putative HYP1  
 Pac-pchgms\_019 18 putative argonaute protein group  
 Pac-pchgms\_019 19 putative protein UXT  
 Pac-pchgms\_019 20 unknown protein  
 Pac-pchgms\_019 21 putative formate dehydrogenase  
 Pac-pchgms\_019 22 putative GAG-POL precursor  
 Pac-pchgms\_019 23 putative polyprotein  
 Pac-pchgms\_019 24 hypothetical protein  
 Pac-pchgms\_019 25 hypothetical protein  
 Pac-pchgms\_019 26 hypothetical protein  
 Pac-pchgms\_019 27 hypothetical protein similar to hCG16339  
 Pac-pchgms\_019 28 hypothetical protein similar to Os12g0257500  
 Pac-pchgms\_019 29 putative integral membrane protein  
 Pac-pchgms\_019 30 putative zinc finger (DHHC type) family protein  
 Pac-pchgms\_019 31 hypothetical protein  
 Pac-pchgms\_019 32 putative CW<sup>14</sup>  
 Pac-pchgms\_019 33 putative DNA binding protein WRKY<sup>2</sup>  
 Pac-pchgms\_019 34 putative aldehyde dehydrogenase  
 Pac-pchgms\_019 35 putative aspartate/glutamate/uridylate kinase family protein  
 Pac-pchgms\_019 36 putative eukaryotic translation initiation factor 5A2

Pac-pchgms\_019 37 putative plastid division regulator MinE  
 Pac-pchgms\_019 38 putative COG1723: Uncharacterized conserved protein  
 Pac-pchgms\_019 39 putative cp protein  
 Pac-pchgms\_019 40 putative CM3 (chorismate mutase 3); chorismate mutase  
 Pac-pchgms\_019 41 hypothetical protein  
 Pac-pchgms\_019 42 putative transferase, transferring glycosyl groups  
 Pac-pchgms\_019 43 putative transferase, transferring glycosyl groups  
 Pac-pchgms\_019 44 hypothetical protein  
 Pac-pchgms\_019 45 putative UNCoordinated family member (unc-95)  
 Pac-pchgms\_019 46 putative pentatricopeptide (PPR) repeat-containing protein  
 Pac-pchgms\_019 47 putative pentatricopeptide (PPR) repeat-containing protein  
 Pac-pchgms\_019 48 putative C-glycosyltransferase  
 Pac-pchgms\_019 49 putative pol polyprotein  
 Pac-pchgms\_019 50 putative IMP dehydrogenase / GMP reductase domain containing protein  
 Pac-pchgms\_019 51 putative ankyrin repeat family protein  
 Pac-pchgms\_019 52 putative ST6-66  
 Pac-pchgms\_019 53 putative remorin family protein  
 Pac-pchgms\_019 54 putative heparanase  
 Pac-pchgms\_019 55 putative pentatricopeptide (PPR) repeat-containing protein  
 Pac-pchgms\_019 56 putative protein binding  
 Pac-pchgms\_020 01 putative FON2 SPARE1  
 Pac-pchgms\_020 02 putative calcineurin-like phosphoesterase family protein  
 Pac-pchgms\_020 03 putative SNAP30 (SOLUBLE N-ETHYLMALIMIDE-SENSITIVE FACTOR ADAPTOR PROTEIN 30); SNAP receptor  
 Pac-pchgms\_020 04 putative RecName: Full=Alpha-1,4 glucan phosphorylase L-2 isozyme, chloroplastic/amyloplastic;  
 Pac-pchgms\_020 05 hypothetical protein  
 Pac-pchgms\_020 06 putative PETHy; ZPT4-2  
 Pac-pchgms\_020 07 hypothetical protein  
 Pac-pchgms\_020 08 putative myb family transcription factor / ELM2 domain-containing protein  
 Pac-pchgms\_020 09 hypothetical protein  
 Pac-pchgms\_020 10 hypothetical protein  
 Pac-pchgms\_020 11 putative F-box family protein  
 Pac-pchgms\_020 12 putative ATP binding / DNA binding / DNA-dependent ATPase  
 Pac-pchgms\_020 13 putative P94  
 Pac-pchgms\_020 14 hypothetical protein  
 Pac-pchgms\_020 15 putative UDP-glucose 6-dehydrogenase  
 Pac-pchgms\_020 16 hypothetical protein  
 Pac-pchgms\_020 17 putative BGLU40 (BETA GLUCOSIDASE 40);  
 Pac-pchgms\_020 18 putative BGLU40 (BETA GLUCOSIDASE 40);  
 Pac-pchgms\_020 19 putative COBW domain containing protein1  
 Pac-pchgms\_020 20 putative NAC domain protein, IPR003441  
 Pac-pchgms\_020 21 hypothetical protein  
 Pac-pchgms\_020 22 putative Gag-Pol polyprotein  
 Pac-pchgms\_020 23 putative NAC domain protein, IPR003441  
 Pac-pchgms\_020 24 putative cell wall hydrolase/autolysin  
 Pac-pchgms\_020 25 putative DRL1 protein  
 Pac-pchgms\_020 26 putative TSD2 (TUMOROUS SHOOT DEVELOPMENT 2); methyltransferase  
 Pac-pchgms\_020 27 putative GagPol3  
 Pac-pchgms\_020 28 putative GagPol3  
 Pac-pchgms\_020 29 putative GAG-POL precursor  
 Pac-pchgms\_020 30 unknown protein  
 Pac-pchgms\_020 31 putative F-box family protein

Pac-pchgms\_020 32 putative F-box family protein  
Pac-pchgms\_020 33 putative F-box family protein  
Pac-pchgms\_020 34 putative TWiK family of potassium channels family member (twk-2)  
Pac-pchgms\_020 35 putative F-box family protein  
Pac-pchgms\_020 36 putative F-box domain-containing protein  
Pac-pchgms\_020 37 hypothetical protein  
Pac-pchgms\_020 38 putative pentatricopeptide (PPR) repeat-containing protein  
Pac-pchgms\_020 39 hypothetical protein  
Pac-pchgms\_020 40 unknown protein  
Pac-pchgms\_020 41 putative RecName: Full=14-3-3-like protein C; AltName: Full=SGF14C  
Pac-pchgms\_020 42 putative mrg-binding protein

## Appendix 2 - Bioinformatic analysis on *ccd4*

### Splicing site prediction

```

SplicePredictor.  Version of July 4, 2007.
Date run: Mon Feb 22 04:34:29 2010

Species:                Medicago truncatula
Model:                  7-class Bayesian
Prediction cutoff (2 ln[BF]):  3.00
Local pruning:         on
Non-canonical sites:   not scored

-----
Sequence 1:  your-sequence, from 1 to 6001.

Potential splice sites

t   q   loc   sequence           P     c     rho  gamma  * P*R*G*   parse
A   <--  156  ccttatgctgcccAGgt  0.980  8.41  0.000  0.000  7 (5 1 1)  -A-
A   <---  540  ctttttttgttgtAGgg  0.996  11.70  0.000  0.016  8 (5 1 2)  I-A-E
A   <--  726  ttgcaaagtatgtAGgt  0.821  3.63  0.000  0.000  7 (5 1 1)  IA-E-EE
A   <--  1146 catatatattttaAGtg  0.923  5.56  0.000  0.000  7 (5 1 1)  IAE-E-EDA
A   <--  1164 ctttgaatgatatAGaa  0.788  3.20  0.000  0.000  7 (5 1 1)  IAEE-E-DADA
D  ----> 1205          tggGTggct  0.890  4.76  0.000  1.885  11 (5 1 5)  IAEEE-D-ADADA
A  <---- 2848 tttcccattttgcAGtg  0.994  10.94  0.000  1.885  11 (5 1 5)  AEEED-A-DADAD
***  ----> 3570          agaGTctaa  0.918  5.39  0.917  1.912  15 (5 5 5)  AEDA-D-ADAD
***A <----- 3773 tacttttgtaacAGgc  0.999  14.40  0.917  1.869  15 (5 5 5)  DAD-A-DAD
D  ----> 5143          gagGTatgc  0.870  4.39  0.768  1.753  15 (5 5 5)  DA-D-AD
A  <----- 5243 tgattaacttttcAGga  0.883  4.61  0.768  1.736  15 (5 5 5)  D-A-D
D  --> 5343          attGTactc  0.854  4.10  0.000  0.000  7 (5 1 1)  -D-

```

### Protein targeting

```

### targetp v1.1 prediction results #####
Number of query sequences: 1

Name                Len     cTP     mTP     SP  other  Loc  RC  TPlen
-----
Sequence            597    0.895  0.051  0.046  0.076  C    1    78
-----
cutoff                0.000  0.000  0.000  0.000

```

```

### chlorop v1.1 prediction results #####
Number of query sequences: 1

Name          Length      Score  cTP      CS-      cTP-
score length
-----
Sequence      597         0.560  Y       5.717      78
-----

```

### Appendix 3 - ccd4 alleles

#### Allele W<sup>1</sup>

```

LOCUS      Pp-ccd4          2849 bp      DNA      linear      PLN 12-JUL-2012
DEFINITION Carotenoid Cleavage Dioxygenase 4 gene.
ACCESSION  Pp-ccd4
VERSION
KEYWORDS
SOURCE     Prunus persica (peach)
ORGANISM   Prunus persica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; fabids; Rosales; Rosaceae; Amygdaloideae; Amygdaleae;
Prunus.
REFERENCE  1 (bases 1 to 2849)
AUTHORS    Adami,M., De Franceschi,P., Brandi,F., Liverani,A., Giovannini,D.,
Rosati,C., Dondini,L. and Tartarini,S.
TITLE      One gene, two colors, three mutations: unraveling the flesh color
determinism in peach
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 2849)
AUTHORS    Adami,M.
TITLE      Direct Submission
JOURNAL    Submitted (11-JUL-2012) Dipartimento di Colture Arboree, Universita
degli Studi di Bologna, Via fanin 46, Bologna, Bologna 40127, Italy
COMMENT    Bankit Comment: ALT EMAIL:ci15399@gmail.com.
Bankit Comment: TOTAL # OF SEQS:3.
Bankit Comment: TOTAL # OF SETS:1.
FEATURES   Location/Qualifiers
source     1..2849
/organism="Prunus persica"
/mol_type="genomic DNA"
/cultivar="White Redhaven"
/db_xref="taxon:3760"
gene      628..2630
/gene="Pp-ccd4"
CDS       join(628..1321,1531..2630)
/gene="Pp-ccd4"
/codon_start=1
/product="Carotenoid Cleavage Dioxygenase"
/translation="MDAFSSSFLSTFPQTQNLSSLSPAIATPKFSISSVRIERPSSPPP
ASKPTSTKAPQPKTPSPPLTTKARDYNNASTFSAKKGTDPPLPAVIFNALDDIINN
FIDPPLRPSVDPKHVLSNMFAPVDELPPTECEIIQGSLPPCLDGAYIRNGPNPQYLPR
GPYHLFDGDGMLHSVRISKGRAVLCSRYVKTYKTIERDAGYPILPSVFSGFNGLTAS
ATRGALSAARVFTGQYNPANGIPLANTSLAFFGNQLYALGESDLPYSLRLTNSGDIQT
LGRHDFDGKLFMSMTAHPKIDPETGEAFAFRYGPLPPFLTYFRFDANGTKQPDVPIFS
MVTPSFLHDFAITKKYALFVDIIGMNPIDMITKGASPVGLDPSKVPRIQVPIPRYAKD
ETEMRWFDPVPGFNIHAINAWDEEDAIVMVAPNILSAEHTMERMDLIHASVEKVRIDL
KTGIVSRQPISTRNLDFAVFNPAYVVGKKNKYVYAAVGDPMPKISGVVKLDVSNVEHKE
CIVASRMFGPGCYGGPEFFVAREPENPEADEDDGYVVVTVYVHDEKAGESSFLVMDAKSP
RLDIVADVRLPRRVPYGFHGLFVKESDLNKL"

```

```

repeat_region 674..689
/satellite="microsatellite:Ppccd4-ssr"
BASE COUNT 825 a 687 c 571 g 766 t
ORIGIN
1 agcgttcgat ttagggttta ggtgtctgat ttccttcgat gataaaaaaa aaatatagat
61 aaaagaaaag gcataacaac tttaaaccag gtccacacag acaaaaacaag ccttgaaaga
121 agtccttgcg toccttaag gaccatctct atgaagagag ccatttatgc catcatggtc
181 tctctctctc tcttctccat gagagagagt tttaaaaagt taaaagagaa agcacttgcc
241 gatagcttga tactaaagat attaaccxaa atactaaaca aattgagggt agaaaaggag
301 caagatgaag gtgacaaggg tgcccctctg caataaatgc attttggtga aacatcaacc
361 tagcgaggat taccagatat aatttttgaa ttgaaatttc tgaaccatcc atttggggaa
421 gcggaagata ttgcagaatg tggccccctc ctctgttctg attgtaacca gaattatitg
481 agagtgtcct cacacagcaa gcaaatatcc caacaccaac atgaagcctt tatttattcc
541 aaccaactga tcccacacct cactaccttt ttttctctct tttgttaatt actttcccat
601 tttgcagtga agggcaatac cagaattatg gatgccttct cttcctcttt cctatccaca
661 tttcccactc aaaaatctctc tctctctcct gccatagcaa cccccaagtt cagcatttcc
721 tctgttagaa ttgaagaaa gccaacaagc ccaccaccag cttcaaaaacc cactccaca
781 aaagcaccac agcctccaaa aactccatcc cgcgccactaa caacaaaagc tgcggtattc
841 aacaatgctt caacattctc tgcggcaaaa aaaggaacag atcctacgct acctgcggtg
901 atcttcaacg ctttggatga catcataaac aacttcatag accctccact gcgccttct
961 gtggaccxaa agcagctcct ctccaacaac tttgctccgg ttgatgagct tcctccgacc
1021 gagtgtgaga tcatacaggg ctccctaccg cegtgcctcg acggtgccta catccgcaat
1081 ggcccgaacc cgagctacct tccgcgtggg cctaccacc tgtttgacgg agacggcatg
1141 cttcactctg ttaggatctc caaaggcctg gccgtgctgt gcagccgcta tgtcaagacc
1201 tacaataaca ccattgagcg tgatgctgcg taccctatc tcccagcgt cttctctggc
1261 ttcaacggcc tcaactgctc cgccacacgt ggcgctctct ccgctgcccg cgtctttaca
1321 ggcaagctct aaaaatcttca ctaatactat ttaaatcaat ttggcaaaa tacacactaa
1381 ccccatgggt ttattaatc actcaattta taagatgcga acaattatac taccaaacta
1441 taaactcgaa gtcggatgca taaatacaat tttttagaat agaatttcc atatgacttt
1501 tcttaataat ttcaataact ttggtaacag gccaaataca tcctgcta ggcattggtc
1561 ttgcaaacac aagtttggtt ttctttggca accaacttta tgcgcttggc gactctgatc
1621 tcccttattc tttgcgtttg acatccaatg gagatatcca aactctggga cgccatgatt
1681 ttgatggcaa gctcttcatg agcatgacg ctcatccaaa gatagaccct gaaacagggg
1741 aggcctttgc cttccgctac ggccccttgc ctccatttct aacatacttt cggtttgatg
1801 caaacggcac aaagcagcca gacgtgccca tattttctat ggtcactcca tctgtcctcc
1861 acgactttgc aatcacaaaa aaatatgcca tctttgttga catacaaat ggcatagaacc
1921 caattgacat gatcacaaaa ggagcatcgc ctggttgctt gacccttcc aagggtccta
1981 gaatcggagt gatcccgcga tacgcgaaa agtgaacgga gatgagggtg tttgatgtgc
2041 ctggttttaa catcatacat gccatcaatg cttgggatga agaggatgct atagtgatgg
2101 ttggcaccaaa cattttgtca gcagagcaca caatggagag aatggatttg atccatgctt
2161 cggttgagaa agttaggatt gatctcaaga cagggtattg gtcaaggcaa cccatctcca
2221 caagaaatct agactttgct gtgttcaatc cggcttatgt ggggaagaag aacaagtatg
2281 tttatgcagc tgtgggtgat ccaatgccta agatatcagg gttggtgaaa ttgatgtgt
2341 ccaatgtgga gcacaaggag tgtatagtgg ctgacaggat gtttgggcca ggggtctatg
2401 ggggtgagcc gttttcgtg gctagagagc ccgagaatcc ggaggcagat gaggacgatg
2461 ggtacgtggt gacgtatggt catgatgaga aggcaggaga atcaagctt ttgggtgatg
2521 atgccaagtc tcctcggctt gatattgtgg cagatgtgag gctgccccgg cgggtgctt
2581 atggcttcca tggactcttt gtgaaggaaa gtgatctcaa caagttgtag tgtcatgtgt
2641 cccattcag caacatagta aacgttacac aaactataaa aataaaaata aaacaccac
2701 cacatatata tactgtcata acccaaac attacaatta caagtttgaa atatatatg
2761 tacatatttt caataaagct atctgcattt tatttactga ttatatttac caccaaaata
2821 aagattatac cccaaactag gtgagtgca
//

```

## AlleleY<sup>1</sup>

LOCUS Pp-ccd4-Y<sup>1</sup> 2860 bp DNA linear PLN 12-JUL-2012  
DEFINITION Carotenoid Cleavage Dioxygenase 4 gene with frameshift caused by STR.  
ACCESSION Pp-ccd4-Y<sup>1</sup>  
VERSION  
KEYWORDS .  
SOURCE Prunus persica (peach)  
ORGANISM Prunus persica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; fabids; Rosales; Rosaceae; Amygdaloideae; Amygdaleae;  
Prunus.  
REFERENCE 1 (bases 1 to 2860)  
AUTHORS Adami,M., De Franceschi,P., Brandi,F., Liverani,A., Giovannini,D.,  
Rosati,C., Dondini,L. and Tartarini,S.  
TITLE One gene, two colors, three mutations: unraveling the flesh color  
determinism in peach  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2860)  
AUTHORS Adami,M.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUL-2012) Dipartimento di Colture Arboree, Universita  
degli Studi di Bologna, Via fanin 46, Bologna, Bologna 40127, Italy  
COMMENT Bankit Comment: ALT EMAIL:cil5399@gmail.com.  
Bankit Comment: TOTAL # OF SEQS:3.  
Bankit Comment: TOTAL # OF SETS:1.  
FEATURES Location/Qualifiers  
source 1..2860  
/organism="Prunus persica"  
/mol\_type="genomic DNA"  
/cultivar="Redhaven"  
/db\_xref="taxon:3760"  
repeat\_region 684..701  
/satellite="microsatellite:Ppccd4-ssr"  
BASE COUNT 831 a 688 c 572 g 769 t  
ORIGIN  
1 aaataagtaa gcgttcgatt tagggtttag gtgtctgatt tccttcgatg ataaaaaaaa  
61 aatatagata aaagaaaagg cataacaact ttaaaccagg tccacacaga caaacaagc  
121 cttgaaagaa gtccttgctt ccctaagg accatctcta tgaagagagc catttatgcc  
181 atcatgggtc ctctctctct cttctccatg agagagagtt ttaaaaagtt aaaagagaaa  
241 gcacttgccg atagcttgat actaaagata ttaacccaaa tactaaacaa attgaggtta  
301 gaaaaggagc aagatgaagg tgacaagggt gccctctgc aataaatgca ttttggtgaa  
361 acatcaacct agcgaggatt accagatata atttttgaat tgaatttct gaaccatcca  
421 tttggggaag cggaagatat tgcagaatgt ggtcccctcc tctgttctga ttgtaaccag  
481 aattatttga gagtgctctc acacagcaag caaatatccc aacaccaaca tgaagccttt  
541 atttattcca accaactgat cccacacctc actacctttt tttctctctt ttgttaatta  
601 ctttccatt ttgcagtga gggcaatacc agaattatgg atgccttctc ttctctttc  
661 ctatccacat ttccactca aaatctctct ctctctctcc tgccatagca accccaagt  
721 tcagcatttc ctctgttaga attgaagaaa gaccatcaag cccaccacca gttcaaaa  
781 ccacctccac aaaagacca cagcctccaa aaactccatc cccgccacta acaaaaaag

841 ctgcgatta caacaatgct tcaacattct ctgcggcaaa aaaaggaaca gatcctacgc  
901 tacctgcggt gatcttcaac gctttggatg acatcataaa caacttcata gaccctccac  
961 tgcgcccttc tgtggacca aagcacgtcc tctccaacaa ctttgctccg gttgatgagc  
1021 ttcctccgac cgagtgtgag atcatacagg gctccctacc gccgtgcctc gacggtgctc  
1081 acatccgcaa tggcccgaac cgcagatacc ttccgcgtgg gccctaccac ctgtttgacg  
1141 gagacggcat gcttactct gttaggatct ccaaaggccg tgccgtgctg tgcagccgct  
1201 atgtcaagac ctacaaatac accattgagc gtgatgctgg ctaccctatt cttcccagcg  
1261 tcttctctgg cttcaacggc ctcactgctc cgcaccacag tggcgctctc tccgctgcc  
1321 gcgtctttac aggcagagtc taaaattttc actaatacta tttaaatcaa tttggcaaaa  
1381 ttacacacta accccatggt tttattaatt cactcaattt ataagatgcy aacaattata  
1441 ctaccaaact ataaactcga agtcggagtc ataaatacaa ttttttagaa tagaaatctc  
1501 catatgactt ttcttaatat tttcaatact tttggttaaca ggccaataca atcctgctaa  
1561 tggcattggt cttgcaaaaca caagtttggc tttctttggc aaccaacttt atgcgcttgg  
1621 cgagtctgat ctcccttatt ctttgcgttt gacatccaat ggagatatcc aaactctggg  
1681 acgccatgat tttgatggca agctcttcat gagcatgacg gctcatccaa agatagacc  
1741 tgaaacaggg gaggcctttg ccttccgcta cggccccttg cctccatttc taacatactt  
1801 tcggtttgat gcaaacggca caaagcagcc agacgtgcc atattttcta tggcactcc  
1861 atcgctcctc cagcactttg caatcacaaa aaaatatgcc atctttgttg acatacaaat  
1921 tggcatgaac ccaattgaca tgatcaccaa aggagcatcg cctgttggct tggacccttc  
1981 caaggtgctc agaatcggag tgatcccgcg atacgcgaaa gatgaaacgg agatgaggtg  
2041 gtttgatgtg cctggtttta acatcataca tgccatcaat gcttgggatg aagaggatgc  
2101 tatagtgatg gtggcaccaa acattttgtc agcagagcac acaatggaga gaatggattt  
2161 gatccatgct tcggttgaga aagttaggat tgatctcaag acagggattg tgtcaaggca  
2221 acccatctcc acaagaaatc tagactttgc tgtgttcaat ccggcttatg tggggaagaa  
2281 gaacaagtat gtttatgcag ctgtgggtga tccaatgcct aagatatcag ggtgggtgaa  
2341 attagatgtg tccaatgtgg agcacaagga gtgtatagtg gctagcagga tgtttgggcc  
2401 aggggtgctat ggggtgagc cgtttttctg ggctagagag cccgagaatc cggaggcaga  
2461 tgaggacgat gggtagctgg tgacgtatgt tcatgatgag aaggcaggag aatcaagctt  
2521 ctgggtgatg gatgccaagt ctccctcggc tgatattgtg gcagatgtga ggctgccccg  
2581 gcgggtgcct tatggcttcc atggactctt tgtgaaggaa agtgatctca acaagttgta  
2641 gtgtcatgtg tccccattca gcaacatagt aaacgttaca caaactataa aaataaaaaat  
2701 aaaacaccca ccacatatat atactgtcat aaccacaaac cattacaatt acaagtttga  
2761 aatataatatt gtacatattt tcaataaagc tatctgcatt ttatttactg attatattta  
2821 ccacaaaaat aaagattata ccccaacta ggtgagtga  
//

## Allele Y<sup>2</sup>

LOCUS Pp-ccd4-Y<sup>2</sup> 9113 bp DNA linear PLN 12-JUL-2012  
DEFINITION Carotenoid Cleavage Dioxygenase 4 gene with transposable element insertion.  
ACCESSION Pp-ccd4-Y<sup>2</sup>  
VERSION  
KEYWORDS .  
SOURCE Prunus persica (peach)  
ORGANISM Prunus persica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; fabids; Rosales; Rosaceae; Amygdaloideae; Amygdaleae; Prunus.  
REFERENCE 1 (bases 1 to 9113)  
AUTHORS Adami,M., De Franceschi,P., Brandi,F., Liverani,A., Giovannini,D., Rosati,C., Dondini,L. and Tartarini,S.  
TITLE One gene, two colors, three mutations: unraveling the flesh color determinism in peach  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 9113)  
AUTHORS Adami,M.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUL-2012) Dipartimento di Colture Arboree, Universita degli Studi di Bologna, Via fanin 46, Bologna, Bologna 40127, Italy  
COMMENT Bankit Comment: ALT EMAIL:cil5399@gmail.com.  
Bankit Comment: TOTAL # OF SEQS:3.  
Bankit Comment: TOTAL # OF SETS:1.  
FEATURES Location/Qualifiers  
source 1..9113  
/organism="Prunus persica"  
/mol\_type="genomic DNA"  
/cultivar="Redhaven"  
/db\_xref="taxon:3760"  
repeat\_region 681..696  
/satellite="microsatellite:Ppccd4-ssr"  
mobile\_element 1496..7758  
/mobile\_element\_type="retrotransposon:Gypsy and Copia like"  
BASE COUNT 2787 a 1863 c 1649 g 2814 t  
ORIGIN  
1 aagtcaagcg ttcgatttag ggtttaggtg tctgatttcc ttcgatgata aaaaaaaaaat  
61 atagataaaa gaaaaggcat aacaacttta aaccagggtc acacagacaa aacaagcctt  
121 gaaagaagtc cttgcgtccc ttaagggacc atctctatga agagagccat ttatgccatc  
181 atggctctct tctctctctt ctccatgaga gagagtttta aaaagttaaa agagaaagca  
241 cttgccgata gcttgatact aaagatatta acccaaatac taaacaaatt gaggttagaa  
301 aaggagcaag atgaaggtga caaggggtgc cctctgcaat aaatgcattt tggtgaaaca  
361 tcaacctagc gaggattacc agatataatt tttgaattga aatttctgaa ccatccattt  
421 ggggaagcgg aagatattgc agaatgtggt cccctcctct gttctgatg taaccagaat  
481 tatttgagag tgtcctcaca cagcaagcaa atatcccaac accaacaatga agcctttatt  
541 tattccaacc aactgatccc acacctcaact accttttttt ctctcttttg ttaattactt  
601 tccatttttg cagtgaaggg caataccaga attatggatg ccttctcttc ctctttccta

661 tccacatttc ccaactcaaaa tctctctctc tctcctgcca tagcaacccc caagttcagc  
721 atttctctctg ttagaattga agaaagacca tcaagcccac caccagcttc aaaaccacc  
781 tccacaaaaag caccacagcc tccaaaaact ccatccccgc cactaacaac aaaagctcgc  
841 gattacaaca atgcttcaac atttctctgcg gcaaaaaaag gaacagatcc tacgctacct  
901 gcggtgatct tcaacgcttt ggatgacatc ataaacaact tcatagaccc tccactgcgc  
961 ccttctgtgg acccaaagca cgtcctctcc aacaactttg ctccggttga tgagcttctc  
1021 cgcaccgagt gtgagatcat acagggtctc ctaccgccgt gcctcgacgg tgcctacatc  
1081 cgcaatggcc cgaaccgcga gtaccttccg cgtgggccct accacctgtt tgacggagac  
1141 ggcatgcttc actctgttag gatctccaaa ggccgtgccc tgctgtgcag ccgctatgtc  
1201 aagacctaca aatacaccat tgagcgtgat gctggctacc ctattcttcc cagcgtcttc  
1261 tctggcttca acggcctcac tgccctccgc acacgtggcg ctctctccgc tgcccgcgtc  
1321 tttacaggca gagtctaaaa ttttcaacta tactatttaa atcaatttgg caaaattaca  
1381 cactaacccc atggttttat taattcactc aatttataag atgcgaacaa ttatactacc  
1441 aaactataaa ctogaagtcg gagtcataaa tacaattttt tagaatagaa atttccatat  
1501 tgttataaaa taacctaaat atgtgcgata attcagagct gcacgtaaag tagatgagac  
1561 acagcattta acgaggttcg gctatgccta cgtcctcgga gagcagcagt agtaactttt  
1621 tcaactatgta aaataatagg gctacaagtt tagtgtttac aatataatgt gctcactgaa  
1681 ttttctctct aggagaattt ctctctgctc tctctttcct ctttcttctc tctcttctt  
1741 tctctttttt cttctcttct tcttttccgt ttctcttctt atttatagge tgaataatc  
1801 actattcatc actattcatc actgttcacc cgtgacagac aaactctatc aagccgcaa  
1861 atgaacagta atgaacagta atgaacagta atgaaaactc tatcaagccg ccaaatgaac  
1921 agtaatgaac agtaatgaat agttagtggg ctccacatgt ttattatttt ttattctttt  
1981 ataacactcc cccttgagga ccactatgct gggtgtatca ttaaagactt gcttgagaa  
2041 aacctagtaa ggtagagaac tgatgaaaag aagagtacaa caatctgtat agcatactc  
2101 tggatgctcc ccctgattaa tatctcccc tgatgtcttc atgactatct tttggattga  
2161 gaatctttcg gagtgagtgg atggtgtagc taccaacaat tcatatagta aatatatttc  
2221 cagttagtta tctctatgta aatcatagaa attttcagag tacgcatatc taataacgct  
2281 tgggctattt ataagttcac ttgaaagaat atgccatcac atggtgttat gcggaaatag  
2341 catcaaatat acctcacatc atcccaaat gatggtgatg gagttgagcc ttatctggaa  
2401 aatatgatgt ccggtccaac ataactccgg aaaatcctta aagtgccaa cggataatcc  
2461 tcataaaaaa gcttcaatac tttcttagta taagcaagta tctcgttggc ataagtctc  
2521 atctttaagt cgagacaaat atttctggaa ttcttaagaa gctttaatgt gttgcaaca  
2581 cattataatc aatatactca ctgaggcaat ttatgtacat taatattgag tacagatttc  
2641 gtgcgtcagg cacatgtcca cagacttgct tcatgcaatt tcaatccttt caaggatttt  
2701 gtttaaaggg attaaactct atgacacatt atatagcttt aaccagcta tttatatac  
2761 tttagggaat cacttcaggt gatatgctct gtgcatttaa taacccttaa agataacatg  
2821 ttggtcttcg taaatgtgca ataaagggc acaattgaat ctataaatat ggagaaaaac  
2881 ccaacattcc ttgtttctgc cagaaacagt gtgtcataca aagtaggtat attcccttgt  
2941 attccgaaca cccaagtata aatttcaata ttaacatctt ttggggttcg tactgtgggt  
3001 ttgttcttcc acgttcattc tcatctataa tggaaattacc tctttccatt ttggccaatg  
3061 atattgtcga catttaataa ttgaacgtgg ttccaatcaa cacagccat tgatgatttg  
3121 agtagctact ccaaaatcaa aaatttgcata ttcatcaatt tatgatccca ccattctcat  
3181 gtgcatgccc atgcatcaat tataagcatt tctttgcttt tgggtaccca agtcacttca  
3241 tggggctttg attatgtgag aatgtggatt ataacctcca ttggagcgtt atttagtagt  
3301 agggcgtgga taatctccat ttagggagtt ggaacattta gaaccatata atctgtcatg  
3361 ctgtaggcat gccacagatt aatacatata tgtcaattaa tttctgggac ttttaacctat  
3421 ataatttgcg acgcattagg cgtgcataat atcaatattg acatgtcaaa ggattgtcca  
3481 tttgggactt caatccacat catttgcctac gcaacaggcg tgcataatata tgatcactgc  
3541 tataccatg ttatgttctt atgggacttt taatctgtgc agtgtattat caatgtatcc  
3601 aacttgcaat ctgtaaaatt tgcattaata attcatggat acatacaagt cattcttttg  
3661 gaacggactt atctccccat ttggttacct ttcaagataa aatatcaaat aggtatataa  
3721 gcataaaata acacaagtat tgcttacatc cttaggtggg agaaactttt ctcaagtatc

3781 attcaagctc atatcagccc agtaaatata atgtagcgct tgatgatcta gttatatoct  
3841 gcaatagcaa aatcacaaac tcttttgccct ctgtcaaaac aaataaccct catagttagg  
3901 attacttcat ggattcattc ttcagatggt cattctaaag aaataaaatc tcttatgaac  
3961 agagagttat aaaaagagaa aaaaatgcaa aataatgtga tattgattgc aagagaaggt  
4021 aagcaatcag ggaaggaagc tggtagggagc agacaataag ctctcatatc tccatagtcta  
4081 gaaggacctc aggagattag agcataaggc cgccttcaca atccctgat cttgcagaaa  
4141 catgatcagg gatagtcttg cttcctgaat ggatgatcag agatagtctt gcttctcggg  
4201 cacattaagt gctcactgat aagaaaaaca agtagatata gcatcactag gtatgaagaa  
4261 aattagatgt cttctgcaga gaaaatttcg ttagtaacat atttatacac aaatacaatg  
4321 aataggtaga accggtgaat ttcaaattaa ccataggaaa atgtagagat tctcgggatg  
4381 ctgtggaaaa agtgcgagat tctcgggatg ctttttgaaa aagtagctcc gtgaaatccg  
4441 taaagcaaga tcggctttga aaataatgct tgaaaacgcc gaaagtgccg acaggcatta  
4501 ttagaggcca cgtgaagttt tggactctag gaaagaaaaa gataaatatg ggattcgaga  
4561 agatattggg gtcctgaaaa acagtagcca tatttctgtc atagatattg cctttgcaaa  
4621 ctgctgttca actcaacttc cttcattttt cgaaccttta gtttttctaa gactttcttc  
4681 gaaaccttct taaaaaaaaa aaaaatggctt cctcttcttc ctgcccacaa tatttcaatt  
4741 taaatgatgc tcccacaaca accagtgcg cccaagtttg gcgtccatcc tttgtatcca  
4801 aaaatcgtca tctcacagtt aatgattctg tgatgatgaa tgatgctact gctgtcacgg  
4861 tagctaggaa tttcattatt ccaatggatg aaatgttgtt gacagggaga tctgaggaag  
4921 aggctattga tgactcaatg gcttttagca ttcagagtgc cgcttctgtt tctaactagg  
4981 ctgatcgttt gcgtgttaga gcaaacgagg ttgagaagct gacaactgaa aattcgtctc  
5041 tccaaagaat gcttcatgag tctcaacagg aggttgagaa acttaaagga gagaataatt  
5101 ccttgttgaa actggtgagt tcgtactctg ttgatacact gagaaggcta gacatgctgc  
5161 aggtctccaa tgaaggatt ttgggagacc acgagaagct catggctaag ctaagaggc  
5221 gccgtcctct tccttcagag gcttcagaaa cataatgtaa tttatagat tttacagggc  
5281 ctgcaccttc attgcaggaa aaaaatata ctgttgatg ttcctgtaat aataataatt  
5341 gcgcaacttc ttaaaacttc acctgtggtt ttacgtctt ttcaaatga cggtttgaa  
5401 ccttattcct tataggttca aataaccaca ttgcctctcc caaatttcat atttcaactg  
5461 atggaacttt tggcctgggc taaacacaaa ctcaaatat atttgccctt tttaaatgga  
5521 tttgaaatat gaattgagta aaagccacga taatatcgca atatagtagt gaatagcatt  
5581 aactactata taccacaat ttcaagttca ggatctctca tatattttga tccatgggtt  
5641 tccgcccag atataacaaa atatgtggg agcctcaatt cattatttga ggtttatatt  
5701 gatattatcc atttcgggt gtattcttaa caaccggaat tcacaaaata tatttcttcc  
5761 ttgaggtgct gattataaca taatcgaact ttattaaatt tatcattttc ttatgccaaa  
5821 gaaatattgt gcgtaccacc atttgcaata atacctcaag ggttgtccat ttaactgtt  
5881 gaactttagg ttctcaacac tggtagattt tgaacttcag gccaaaatca catattctca  
5941 tggtagggac attcaatccc cttagatttt gaacttcggg ccaaatcac atattctcat  
6001 ggtatggaca tttttacaat tttctgtaca tatttctgga cttcaagccc ttacatattt  
6061 gtccatattt tgaggaactt ctggcatctc atttaatttc tcatccatga gtttaaggaa  
6121 ctgcaggttc ccttttgat atagtgcagg tttatccaaa atgggttaata tttatgcata  
6181 cgtcactatt catgtgaata gtactgttca tgaatattta tgcatacgtc actattcagg  
6241 tgaatagtag tgttcatgaa tatttatgca tatgtcacta ttcattgtga tagtactgtt  
6301 catcaagtca tctattcatg tgcacagtag atctgccagt acagttatca tccatggctc  
6361 atatttatgc atacatcact attcatgtga atagtactat tcatcaaagt catgaatagc  
6421 tatctattca tctgccagta caattatcat ccatgtgtac ggcactatga accaatacgg  
6481 tactgttaca tcattaagga actctagggt cttatttaca tgtcatggat caaggacctt  
6541 taagtccaat cacatgttta caaatacagt accggagaaa ctgccagctc tcattttaatt  
6601 atcatcatca aggatcttcc agtcctgatg taattgtatg atgaggatca aggaacttct  
6661 ggtcctgatc tgcttactgt aaaaagaact catcatacag cacatttagt ccataaaata  
6721 aattgctggg aaataaatta ctggtatgga cgataaacc gcaccatata tttaaatagc  
6781 agtaaaagtc atgaaataat ttgctggtaa ataaattact agtatggagc ataaaccgc  
6841 accatacatt taaataaatg taaatgtgcg gtaaaataaa tgtgcttcta tgggcaactaa

6901 ttctgctcca tacatttaaa taaaagtaaa ggcgtggacg ataaaccgcg accacccttt  
6961 taaataaatt cataaagtaa atttcataaa gtaaagtgtt agtattagta acggtctccc  
7021 actgataata tgtttagtat taccaagggt ccatttttgt taatggttag taatagaaaa  
7081 gcagataaaa atagtatatt atattacat caacttttca tatatataat atatatatat  
7141 attatactga tatattagtg gaaagctaga tgcgttggct tttcttcttt acatcaacat  
7201 aatggtttagt agtataaata atagtgcagc acaaaaatta tacctgagag cttctcgtgc  
7261 tgataacgtg ttataaaata acctaaatat gtgcgataat tcagagctgc acgtaaagta  
7321 gatgagacac agcatttaac gaggttcggc tatgcctacg tcctcggaga gcagcagtag  
7381 taactttttc actatgtaaa ataatagggc tacaagttta gtgtttacaa tatatgtggc  
7441 tcaactgaatt ttctctctag gagaatttct ctctgctctc tctttcctct ttcttccttc  
7501 tcttcctttc tcttttttct tctcttcttc ttttccgttt ctcttcttat ttataggctg  
7561 aaataatcac tattcatcac tattcatcac tgttcaaccg tgacagacaa actctatcaa  
7621 gccgccaat gaacagtaat gaacagtaat gaacagtaat gaaaactcta tcaagccgcc  
7681 aatgaacag taatgaacag taatgaatag ttagtgggct ccacatgttt attatTTTTT  
7741 attcttttat aacacatatg acttttctta atatTTTcaa tacttttggg aacaggccaa  
7801 tacaatcctg ctaatggcat tggctctgca aacacaagtt tggcttttct tggcaaccaa  
7861 ctttatgctc ttggcgagtc tgatctccct tattctttgc gtttgacatc caatggagat  
7921 atccaaactc tgggacgcca tgattttgat ggcaagctct tcatgagcat gacggctcat  
7981 ccaaagatag accctgaaac aggggagggc tttgccttcc gctacggccc cttgcctcca  
8041 tttctaacat actttcgggt tgatgcaaac ggcacaaagc agccagacgt gccatattt  
8101 tctatggtca ctccatcggt cctccacgac tttgcaatca caaaaaaata tgccatcttt  
8161 gttgacatac aaattggcat gaaccaatt gacatgatca ccaaaggagc atcgctgtt  
8221 ggcttgacc cttccaaggt gcctagaatc ggagtgatcc cgcgatacgc gaaagatgaa  
8281 acggagatga ggtggtttga tgtgcctggt tttaacatca tacatgccat caatgcttgg  
8341 gatgaagagg atgctatagt gatggtggca ccaaactttt tgtcagcaga gcacacaatg  
8401 gagagaatgg atttgatcca tgcttcgggt gagaaaagtt ggattgatct caagacaggg  
8461 attgtgtcaa ggcaaccat ctccacaaga aatctagact ttgctgtgtt caatccggct  
8521 tatgtgggga agaagaacaa gtatgtttat gcagctgtgg gtgatccaat gcctaagata  
8581 tcaggggtgg tgaattaga tgtgtccaat gtggagcaca aggagtgtat agtggctagc  
8641 aggatgtttg ggccagggtg ctatgggggt gagccgtttt tcgtggctag agagcccgag  
8701 aatccggagg cagatgagga cgatgggtac gtggtgacgt atgttcatga tgagaaggca  
8761 ggagaatcaa gcttcttggg gatggatgcc aagtctctc ggcttgatat tgtggcagat  
8821 gtgaggctgc cccggcgggt gccttatggc ttccatggac tctttgtgaa ggaaagtgat  
8881 ctcaacaagt tgtagtgtca tgtgtcccca ttcagcaaca tagtaaactg tacacaaact  
8941 ataaaaataa aaataaaaca cccaccacat atatatactg tcataaccac aaaccattac  
9001 aattacaagt ttgaaatata tattgtacat attttcaata aagctatctg cattttatTT  
9061 actgattata tttaccacca aaataaagat tatacccaa actaggtgag tgc

//

## Allele Y<sup>3</sup>

LOCUS BSeq#1 2154 bp DNA linear PLN 07-NOV-2012  
DEFINITION Prunus persica.  
ACCESSION BSeq#1  
VERSION  
KEYWORDS .  
SOURCE Prunus persica (peach)  
ORGANISM Prunus persica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; fabids; Rosales; Rosaceae; Amygdaloideae; Amygdaleae;  
Prunus.  
REFERENCE 1 (bases 1 to 2154)  
AUTHORS Adami,M., De Franceschi,P., Brandi,F., Liverani,A., Giovannini,D.,  
Rosati,C., Dondini,L. and Tartarini,S.  
TITLE One gene, two colors, three mutations: unraveling the flesh color##  
determinism in peach  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2154)  
AUTHORS Adami,M., De Franceschi,P., Brandi,F., Liverani,A., Giovannini,D.,  
Rosati,C., Dondini,L. and Tartarini,S.  
TITLE Direct Submission  
JOURNAL Submitted (07-NOV-2012) Dipartimento di Colture Arboree, Universita  
degli Studi di Bologna, Via fanin 46, Bologna, Bologna 40127, Italy  
COMMENT Bankit Comment: ALT EMAIL:ci15399@gmail.com.  
Bankit Comment: TOTAL # OF SEQS:1.

##Genome-Assembly-Data-START##

Sequencing Technology :: Sanger dideoxy sequencing

##Genome-Assembly-Data-END##

FEATURES Location/Qualifiers

source 1..2154  
/organism="Prunus persica"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3760"  
gene join(1..696,906..2003)  
/gene="Pp-ccd4"  
/allele="Y3"  
/note="A/T transversion at position 1584 causes premature  
stop codon"  
CDS join(1..696,906..1730)  
/gene="Pp-ccd4"  
/allele="Y3"  
/codon\_start=1  
/translation="MDFSSSFLSTFPQTQNLSPAIATPKFSISSVRIERPSSPPP  
ASKPTSTKAPQPPKTPSPPLTTKARDYNNASTFSAAKKGTDPITLPAVIFNALDDIINN  
FIDPPLRPSVDPKHVLSNNFAPVDELPPTECEI IQGSLPPCLDGAYIRNGPNPQYLPR  
GPYHLFDGDMGLHSVRISKGRAVLCSTRYVKTYKTIERDAGYPILPVSFSGFNGLTAS  
ATRGALSAARVFTGQYNPANGIIGLANTSLAFFGNQLYALGESDLPYSLRLTNSNGDIQT  
LGRHDFDGLFMSMTAHPKIDPETGEAFAFRYGPLPPFLTYFRFDANGTKQPDVPIFS  
MVTPSFLHDFATTKKYAIFVDIQIGMNPIDMITKGASPVGLDPSKVPRIGVIPRYAKD  
ETEMRWFDVPGFNIHAINAWDEEDAIVMVAPNILSAEHTMERMDLIHASVEKVRIDL  
KTGIVSRQPISTRNLDFAVFNPAVYVGKKNKYVYAAVGDPMPKISGVVKLDVSNVEH"  
repeat\_region 49..64  
/satellite="microsatellite:Ppccd4-ssr"  
BASE COUNT 591 a 544 c 451 g 568 t  
ORIGIN

```
1 atggatgcct tctcttctc tttcctatcc acatttccca ctcaaaatct ctctctctct
61 cctgccatag caacccccaa gttcagcatt tcctctgtta gaattgaaga aagaccatca
121 agcccaccac cagcttcaaa acccacctcc acaaaagcac cacagcctcc aaaaactcca
181 tccccgccac taacaacaaa agctcgcgat tacaacaatg cttcaacatt ctctcggcga
241 aaaaaggaa cagatcctac gctacctgog gtgatcttca acgctttgga tgacatcata
301 aacaacttca tagaccctcc actgcgcctt tctgtggacc caaagcacgt cctctccaac
361 aactttgctc cggttgatga gttctctccg accgagtgtg agatcataca gggctcccta
421 cgcctgctgc tcgacggtgc ctacatccgc aatggccgca acccgagta cttcgcgt
481 gggccctacc acctgtttga cggagacggc atgcttcaat ctgttaggat ctccaaaggc
541 cgtgccgtgc tgtgcagcgc ctatgtcaag acctacaat acaccattga gcgtgatgct
601 ggctacceta ttcttcccag cgtcttctct ggcttcaacg gcctcactgc ctccgccaca
661 cgtggcgctc tctccgctgc cgcgctcttt acaggcagag tctaaaatct tcaactaatac
721 tatttaaatc aatttggcaa aattacacac taaccccatg gttttattaa ttcactcaat
781 ttataagatg cgaacaatta tactacaaa ctataaactc gaagtcggag tcataaatac
841 aattttttag aatagaaatt tccatgatgc ttttcttaat attttcaata cttttggtaa
901 caggccaata caatcgtct aatggcattg gtcttgcaaa cacaagtttg gctttctttg
961 gcaaccaact ttatgcgctt ggcgagtctg atctccctta ttctttgcgt ttgacatcca
```

```
1021 atggagatat ccaaactctg ggacgccatg attttgatgg caagctcttc atgagcatga
1081 cggctcatcc aaagatagac cctgaaacag gggaggcctt tgccttccgc tacggcccct
1141 tgcctccatt tctaacatac tttcgggttg atgcaaacgg cacaaagcag ccagacgtgc
1201 ccataatctt tatggctact ccacgttcc tccacgactt tgcaatcaca aaaaaatag
1261 ccatctttgt tgacatacaa attggcatga acccaattga catgatcacc aaaggagcat
1321 cgcctgttgg cttggaccct tccaaggtgc ctagaatcgg agtgatcccg cgatacgcga
1381 aagatgaaac ggagatgagc tggtttgatg tgcctggttt taacatcata catgccatca
1441 atgcttggga tgaagaggat gctatagtga tgggtggcacc aacattttg tcagcagagc
1501 acacaatgga gagaatggat ttgatccatg cttcggttga gaaagttagg attgatctca
1561 agacagggat tgtgtcaagg caaccatct ccacaagaaa tctagacttt gctgtgttca
1621 atccggctta tgtggggaag aagaacaagt atgtttatgc agctgtgggt gatccaatgc
1681 ctaagatata aggggtgggt aaattagatg tgtccaatgt ggagcactag gagtgtatag
1741 tggctagcag gatgtttggg ccaggggtgct atgggggtga gccgttttc gtggctagag
1801 agcccagaaa tccggaggca gatgaggacg atgggtacgt ggtgacgtat gttcatgatg
1861 agaaggcagc agaatcaagc ttcttgggta tggatgcaa gtctcctcgg cttgatattg
1921 tggcagatgt gaggctgccc cggcgggtgc cttatggctt ccatggactc tttgtgaagg
1981 aaagtgatct caacaagttg tagtgtcatg tgtcccatt cagcaacata gtaaacgtta
2041 cacaaactat aaaaaataaa ataaaacacc caccacatat atatactgtc ataaccacaa
2101 accattacaa ttacaagttt gaaatatata ttgtacatat ttccaataaa gcta
//
```

# Bibliography

**Agarwal M, Shrivastava N, Padh H.** 2008. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* **27**: 617–631.

**Ahrazem O, Trapero A, Gómez MD, Rubio-Moraga A, Gómez-Gómez L, Gomez MD, Gomez-Gomez L.** 2010. Genomic analysis and gene structure of the plant carotenoid dioxygenase 4 family: a deeper study in *Crocus sativus* and its allies. *Genomics* **96**: 239–50.

**Aranzana MJ, Abbassi E-K, Howad W, Arús P.** 2010. Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC genetics* **11**: 69.

**Aranzana MJ, Pineda A, Cosson P, Dirlewanger E, Ascasibar J, Cipriani G, Ryder CD, Testolin R, Abbott A, King GJ, et al.** 2003. A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **106**: 819–825.

**Arumuganathan K, Earle EDE.** 1991. Nuclear DNA content of some important plant species. *Plant molecular biology reporter* **9**: 208–218.

**Arús P, Verde I, Sosinski B, Zhebentyayeva T, Abbott AG.** 2012. The peach genome. *Tree Genetics & Genomes* **8**: 531–547.

**Auldridge ME, McCarty DR, Klee HJ.** 2006. Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Current opinion in plant biology* **9**: 315–21.

**Bailey J, French A.** 1949. The inheritance of certain fruit and foliage characters in the peach.

**Bassi D, Monet R.** 2008. Botany and taxonomy. In: Bassi D, D. L, eds. *The peach: botany, production and uses*. Wallingford: CAB International, 1–36.

**Beckman T, Alcazar J.** 2005. Evidence for qualitative suppression of red skin color in peach. *HortScience* **40**: 523–524.

**Beckman T, Sherman W.** 2003. Probable qualitative inheritance of full red skin color in peach. *HortScience* **38**: 1184–1185.

**Bian W, Barsan C, Egea I, Purgatto E, Chervin C, Zouine M, Latché A, Bouzayen M, Pech J-C.** 2011. Metabolic and molecular events occurring during chromoplast biogenesis. *Journal of Botany* **2011**: 1–13.

**Brandi F, Bar E, Mourgues F, Horváth G, Turcsi E, Giuliano G, Liverani A, Tartarini S, Lewinsohn E, Rosati C.** 2011. 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 biology* **11**: 24.

**Brandi F,** 2010. Studi biochimici, molecolari e genetici dell’accumulo dei carotenoidi in genotipi di pesco a polpa bianca e gialla. PhD Thesis, University of Bologna.

- Byrne D, Nikolic A, Burns E.** 1991. Variability in sugars, acids, firmness, and color characteristics of 12 peach genotypes. *Journal of the American Society for Horticultural Science* **116**: 1004–1006.
- Cao K, Wang L, Zhu G, Fang W, Chen C, Zhao P.** 2010. Study on molecular marker linked to flesh color characteristics using SSR in peach and its confirmation in different varieties. *Journal of Fruit Science* **6**: 7.
- Cazzonelli CI, Roberts AC, Carmody ME, Pogson BJ.** 2010. Transcriptional control of SET DOMAIN GROUP 8 and CAROTENOID ISOMERASE during Arabidopsis development. *Molecular plant* **3**: 174–91.
- Chagné D, Gasic K, Crowhurst RRN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EHA, Gardiner SE, Korban SS.** 2008. Development of a set of SNP markers present in expressed genes of the apple. *Genomics* **92**: 353–8.
- Chaparro JXJX, Werner DJDDJ, O'Malley D, Sederoff RR.** 1994. Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. *Theoretical and Applied Genetics* **87**: 805–815.
- Chu Y, Su X, Huang Q, Zhang X.** 2009. Patterns of DNA sequence variation at candidate gene loci in black poplar (*Populus nigra* L.) as revealed by single nucleotide polymorphisms. *Genetica* **137**: 141–50.
- Connors CH.** 1920. Some notes on the inheritance of unit characters in the peach. *Proceedings American Society of Horticultural Science* **16**: 24–36.
- Cunningham FX, Gantt E.** 1998. Genes and enzymes of carotenoid biosynthesis in plants. *Annual review of plant physiology and plant molecular biology* **49**: 557–583.
- Cunningham FX, Gantt E.** 2001. One ring or two? Determination of ring number in carotenoids by lycopene epsilon-cyclases. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 2905–10.
- Cunningham FX, Sun Z, Chamovitz D, Hirschberg J, Grantt E.** 1996. Functional analysis of the beta and epsilon lycopene cyclase enzymes of *Arabidopsis* reveals a mechanism for control of cyclic carotenoid formation. *Plant Cell* **8**: 1613–1626.
- David H. Byrne, Maria Bassols Raseira, Daniele Bassi, Maria Claudia Piagnani, Ksenija Gasic, Gregory L. Reighard, María Angeles Moreno and SP.** 2012. Peach. In: Badenes ML, Byrne DH, eds. *Fruit Breeding, Handbook of Plant Breeding* 8. Boston, MA: Springer US, 505–569.
- Decroocq V, Favé MG, Hagen L, Bordenave L, Decroocq S.** 2003. Development and transferability of apricot and grape EST microsatellite markers across taxa. *TAG. Theoretical and Applied Genetics* **106**: 912–22.
- Dermen H.** 1960. Nature of plant sports. *American Horticultural Magazine* **39**: 123–73.
- Dermen H, Stewart R.** 1973. Ontogenetic study of floral organs of peach (*Prunus persica*) utilizing cytochimeral plants. *American Journal of Botany* **60**: 283–291.

- Diretto G, Welsch R, Tavazza R, Mourgues F, Pizzichini D, Beyer P, Giuliano G.** 2007. Silencing of beta-carotene hydroxylase increases total carotenoid and beta-carotene levels in potato tubers. *BMC plant biology* **7**: 11.
- Dirlewanger E, Cosson P, Tavaud M, Aranzana J, Poizat C, Zanetto A, Arús P, Laigret F.** 2002. Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **105**: 127–138.
- Dirlewanger E, Graziano E, Joobeur T, Garriga-Calderé F, Cosson P, Howad W, Arús P, Garriga-Calderé F, Arus P.** 2004. Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 9891–9896.
- Dominguez I, Graziano E, Gebhardt C, Barakat A, Berry S, Arús P, Delseny M, Barnes S.** 2003. Plant genome archaeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species. *Plant biotechnology journal* **1**: 91–9.
- Eduardo I, Pacheco I, Chietera G, Bassi D, Pozzi C, Vecchietti A, Rossini L.** 2010. QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. *Tree Genetics & Genomes* **7**: 323–335.
- Etienne C, Rothan C, Moing A, Plomion C, Bodénès C, Svanella-Dumas L, Cosson P, Pronier V, Monet R, Dirlewanger E.** 2002. Candidate genes and QTLs for sugar and organic acid content in peach [*Prunus persica* (L.) Batsch]. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **105**: 145–159.
- Farré G, Sanahuja G, Naqvi S, Bai C, Capell T, Zhu C, Christou P.** 2010. Travel advice on the road to carotenoids in plants. *Plant Science* **179**: 28–48.
- Faust M, Timon B.** 1995. Origin and Dissemination of Peach. In: Jules J, ed. Horticultural Reviews. New York: John Wiley & Sons, Inc., 331–379.
- Folta KM, Gardiner SE, Pozzi C, Vecchietti A.** 2009. 11 . Peach Structural Genomics (RA Jorgensen, Ed.). *Genetics and Genomics of Rosaceae* **6**: 235–257.
- Fray RG, Grierson D.** 1993. Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant Molecular Biology* **22**: 589–602.
- Garvin MR, Saitoh K, Gharrett AJ.** 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular ecology resources* **10**: 915–34.
- Giuliano G, Bartley GE, Scolnik PA.** 1993. Regulation of carotenoid biosynthesis during tomato development. *Plant Cell* **5**: 379–387.
- Gómez-Gómez L, Rubio-Moraga Á, Ahrazem O.** 2010. Understanding carotenoid metabolism in affron stigmas: unravelling aroma and colour formation. *Functional Plant Science and Biotechnology* **4**: 56–63.
- Gradziel T, Beres W.** 1993. Semidwarf growth habit in clingstone peach with desirable tree and fruit qualities. *HortScience* **28**: 1045–1047.

- Hayden MJ, Tabone T, Mather DE.** 2009. Development and assessment of simple PCR markers for SNP genotyping in barley. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **119**: 939–51.
- Howad W, Yamamoto T, Dirlwanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG, Arús P, et al.** 2005. Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. *Genetics* **171**: 1305–1309.
- Huang F-CC, Molnár P, Schwab W, Molnar P.** 2009. Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. *Journal of experimental botany* **60**: 3011–22.
- Illa E, Sargent DJ, Lopez Girona E, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, Van der Knaap E, Iezzoni A, et al.** 2011. Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. *BMC evolutionary biology* **11**: 9.
- Joobeur T, Viruel MA, De Vicente MC, Jáuregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Batlle I, et al.** 1998. Construction of a saturated linkage map for *Prunus* using an almond×peach F<sub>2</sub> progeny. *TAG Theoretical and Applied Genetics* **97**: 1034–1041.
- Klee HJ, Giovannoni JJ.** 2011. Genetics and control of tomato fruit ripening and quality attributes. *Annual review of genetics* **45**: 41–59.
- Kohany O, Gentles AJ, Hankus L, Jurka J.** 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **7**: 474.
- Kumar P, Gupta V, Misra A, Modi D, Pandey B.** 2009. Potential of Molecular Markers in Plant Biotechnology. *Plant Omics Journal* **2**: 141.
- Kuzin AB, Lyubomirskaya N V, Khudaibergenova BM, Ilyin Y V, Kim AI.** 1994. Precise excision of the retrotransposon *gypsy* from the *forked* and *cut* loci in a genetically unstable *D. melanogaster* strain. *Nucleic Acids Research* **22**: 4641–4645.
- Lesley J.** 1940. A genetic study of saucer fruit shape and other characters in the peach. *Proc Am Soc Hortic Sci.*
- Li J, Dudas B, Webster MA, Cook HE, Davies BH, Gilmartin PM.** 2010. *Hose in Hose*, an S locus-linked mutant of *i*, is caused by an unstable mutation at the *Globosa* locus. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 5664–8.
- Li Y-C, Korol AB, Fahima T, Nevo E.** 2004. Microsatellites within genes: structure, function, and evolution. *Molecular biology and evolution* **21**: 991–1007.
- Litt M, Luty J.** 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American journal of human genetics* **44**: 397–401.

- Lu S, Li L.** 2008. Carotenoid Metabolism: Biosynthesis, Regulation, and Beyond. *Journal of Integrative Plant Biology* **50**: 778–785.
- Mignani I, Ortugno C, Bassi D.** 2005. Biochemical parameters for the evaluation of different peach flesh types. *VI International Peach Symposium* **713**.
- Mnejja M, Garcia-Mas J, Audergon J-MJ, Arús P.** 2010. Prunus microsatellite marker transferability across rosaceous crops. *Tree Genetics & Genomes* **6**: 689–700.
- Moehs CP, Tian L, Osteryoung KW, Dellapenna D.** 2001. Analysis of carotenoid biosynthetic gene expression during marigold petal development. *Plant molecular biology* **45**: 281–93.
- Morgante M, Hanafey M, Powell W.** 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature genetics* **30**: 194–200.
- Morgutti S, Negrini N, Nocito FF, Ghiani A, Bassi D, Cocucci M.** 2006. Changes in endopolygalacturonase levels and characterization of a putative endo-PG gene during fruit softening in peach genotypes with nonmelting and melting flesh fruit phenotypes. *The New phytologist* **171**: 315–28.
- Morrison DM.** 1990. Color and beta-carotene in six genotypes of peach.
- Nefedova LN, Ljubomirskaya N V., Ilyin Y V., Kim a. I.** 2006. Precise excision of long terminal repeats of the gypsy (mdg4) retrotransposon of *Drosophila melanogaster* detected in *Escherichia coli* cells is explained by its integrase function. *Russian Journal of Genetics* **42**: 1398–1404.
- Niyogi KK, Bjorkman O, Grossman AR.** 1997. The roles of specific xanthophylls in photoprotection. *PNAS* **94**: 14162–14167.
- Ogundiwin EA, Peace CP, Gradziel TM, Parfitt DE, Bliss FA, Crisosto CH.** 2009. A fruit quality gene map of Prunus. *BMC genomics* **10**: 587.
- Ohmiya A, Kishimoto S, Aida R, Yoshioka S, Sumitomo K.** 2006. Carotenoid cleavage dioxygenase (*CmCCD4a*) contributes to white color formation in chrysanthemum petals. *Plant physiology* **142**: 1193–201.
- Van Ooijen JW, Voorrips RE.** 2001. JoinMap 3.0, Software for the calculation of genetic linkage maps (JW Van Ooijen and RE Voorrips, Eds.). *Plant Research Plant Research International*.
- Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, et al.** 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nature biotechnology* **23**: 482–7.
- Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD.** 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* **127**: 181–97.
- Peace C, Ahmad R.** 2004. The use of molecular genetics to improve peach and nectarine post-storage quality. V International Postharvest Symposium.

- Pirazzoli C.** 2008. Il Mercato. In: Angelini R, ed. Bologna: Script, 19–27.
- Pleines T, Jakob SS, Blattner FR.** 2008. Application of non-coding DNA regions in intraspecific analyses. *Plant Systematics and Evolution* **282**: 281–294.
- Pogson B, McDonald KA, Truong M, Britton G, DellaPenna D.** 1996. *Arabidopsis* carotenoid mutants demonstrate that lutein is not essential for photosynthesis in higher plants. *The Plant cell* **8**: 1627–39.
- Quilot B, Kervella J.** 2005. Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. *Journal of Experimental ...*
- Rafalski JA.** 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science* **162**: 5.
- Rubio A, Rambla JL, Santaella M, Gómez MD, Orzaez D, Granell A, Gómez-Gómez L.** 2008. Cytosolic and plastoglobule-targeted carotenoid dioxygenases from *Crocus sativus* are both involved in beta-ionone release. *The Journal of biological chemistry* **283**: 24816–25.
- Sanzol J.** 2009. Pistil-function breakdown in a new S-allele of European pear, S21\*, confers self-compatibility. *Plant cell reports* **28**: 457–67.
- Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen J-E, Weigel D, Andersen SU.** 2009. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature methods* **6**: 550–1.
- Schuster SC.** 2008. Next-generation sequencing transforms today's biology. *Nature methods* **5**: 16–8.
- Scott D, Cullinan F.** 1942. The inheritance of wavy-leaf character in the peach. *Journal of Heredity* **33**: 293–295.
- Shewmaker CK, Sheehy JA, Daley M, Colburn S, Ke DY.** 1999. Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. *Plant Journal* **20**: 401–412.
- Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, et al.** 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nature genetics* **43**: 109–16.
- Sosinski B, Gannavarapu M, Hager LD, Beck LE, King GJJ, Ryder CD, Rajapakse S, Baird W V., Ballard RE, Abbott a. G.** 2000. Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. *TAG. Theoretical and Applied Genetics* **101**: 421–428.
- Szymkowiak EJ, Sussex IM.** 1996. What Chimeras Can Tell Us About Plant Development. *Annual review of plant physiology and plant molecular biology* **47**: 351–376.
- Toledo-Ortiz G, Huq E, Rodríguez-Concepción M.** 2010. Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 11626–31.

- Toyama T.** 1974. Haploidy in peach. *HortScience* **9**: 187–188.
- Varshney RK, Nayak SN, May GD, Jackson S a.** 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in biotechnology* **27**: 522–30.
- Verde I, Abbott AG, Salamini F, Schmutz J, Sosinski B, Morgante M, Rokhskar DS, Lucas S, Cattonaro F, Grimwood J, et al.** The draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution.
- Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, Micheletti D, Rosyara UR, Cattonaro F, Vendramin E, et al.** 2012. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. (R Wu, Ed.). *PloS one* **7**: e35668.
- Vizzotto M, Cisneros-Zevallos L, Byrne D, Okie W, Ramming D.** 2006. Total Phenolic, Carotenoids, and Anthocyanin Content and Antioxidant Activity of Peach and Plum Genotypes. *Acta Horticulturae* **713**: 453–455.
- Walter MH, Strack D.** 2011. Carotenoids and their cleavage products: biosynthesis and functions. *Natural Product Reports* **28**: 663–92.
- Wang ML, Barkley NA, Jenkins TM.** 2009. Microsatellite markers in plants and insects. Part I: Applications of biotechnology. *Genes, genomes and genomics* **3**: 54–67.
- Warburton ML, Becerra-Velásquez VL, Goffreda JC, Bliss FA.** 1996. Utility of RAPD markers in identifying genetic linkages to genes of economic interest in peach. *Theoretical and Applied Genetics* **93-93**: 920–925.
- Werner D, Creller M.** 1997. Genetic studies in peach: inheritance of sweet kernel and male sterility. *Journal of the American Society for Horticultural Science* **122**: 215–217.
- Werner D, Creller M, Chaparro J.** 1998. Inheritance of the blood-flesh trait in peach. *HortScience* **33**: 1243–1246.
- Williamson JD, Peace CP, Bliss FA, Garner DT, Crisosto CH.** 2006. Evidence for a single locus controlling flesh color, senescent leaf color, and hypanthium color in peach. *Journal of the American Society for Horticultural Science* **131**: 256–260.
- Yan J, Kandianis CB, Harjes CE, Bai L, Kim E-H, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, et al.** 2010. Rare genetic variation at *Zea mays* crtRB1 increases beta-carotene in maize grain. *Nature genetics* **42**: 322–7.
- Yoshida M, Yamane K, Ijiro Y.** 2000. Studies on ornamental peach cultivars. *Bulletin of the College of Agriculture-Utsunomiya University* **17**: 1–14.
- Yoshioka S, Aida R, Yamamizo C, Shibata M, Ohmiya A.** 2011. The *carotenoid cleavage dioxygenase 4* (*CmCCD4a*) gene family encodes a key regulator of petal color mutation in chrysanthemum. *Euphytica* **184**: 377–387.

**Ytterberg A, Peltier J, Wijk K Van.** 2006. Protein profiling of plastoglobules in chloroplasts and chromoplasts. A surprising site for differential accumulation of metabolic enzymes. *Plant Physiology* **140**: 984–997.

**Zamboni A, Pierantoni L, De Franceschi P.** 2008. Total RNA extraction from strawberry tree ( *Arbutus unedo* ) and several other woody-plants. *iForest - Biogeosciences and Forestry* **1**: 122–125.

**Zhou X, Van Eck J, Li L.** 2008. Use of the cauliflower *Or* gene for improving crop nutritional quality. *Biotechnology annual review* **14**: 171–90.

# Summary

Introduction .....	1
Peach systematics .....	3
The origin of modern Peach.....	4
Peach economics .....	5
Mutations and chimerism.....	6
Phenotypic traits.....	9
Fruit traits .....	10
Peach flesh color.....	12
Carotenoids.....	14
Main roles of carotenoids.....	15
Carotenoid biosynthesis pathway .....	16
Carotenoid accumulation .....	17
Differential carotenoid accumulation and gene expression in peach.....	19
Molecular markers and mapping.....	21
SSR markers .....	21
SNP.....	22
Mapping.....	23
The peach genome .....	27
Aim of the work.....	33
Material and Methods.....	35
Plant material.....	37
DNA extraction.....	40
DNA extraction from leaf.....	40
DNA extraction from embryos.....	41
PCR .....	42
Standard amplifications.....	42
Insertion detection (Three primer system) .....	42

Amplification of long templates and Hi-Fidelity PCR .....	43
Temperature switch PCR (TSP).....	43
Primer design .....	44
Agarose gel electrophoresis.....	45
Polyacrylamide gel electrophoresis .....	46
Cloning.....	48
Ligase.....	48
A-tailing .....	48
Competent cells .....	49
Transformation .....	49
Growth medium.....	50
Plasmid extraction.....	51
Expression vector construction .....	52
RNA extraction .....	53
DNase treatment.....	54
Retrotranscription.....	54
Mapping .....	55
In-silico analysis.....	56
Phylogenetic analysis .....	56
Rice GAAS analysis .....	57
Results.....	59
Molecular characterization of mutant pairs .....	61
Identification in-silico of genes related to white-yellow phenotype within the Y locus .....	62
Phylogenetic analysis of carotenoid dioxygenases .....	63
Peach and apple CCD4 comparison.....	65
Development and characterization of ccd4-SSR marker.....	67
Cloning and characterization of the Redhaven (TC) <sub>8</sub> allele.....	70

Analysis of the chimerical fruit tissues of White Redhaven .....	73
Characterization of Caldesi 2000/Cristina mutant system .....	73
Mapping and validation of <i>ccd4</i> co-segregation with Y-locus .....	74
Analysis of <i>ccd4</i> in peach germplasm .....	76
Expression and functional verification.....	81
Discussion.....	83
Conclusions .....	88
Additional material.....	89
Appendix 1 - RiceGAAS Predicted function.....	91
Appendix 2 - Bioinformatic analysis on <i>ccd4</i> .....	113
Splicing site prediction.....	113
Protein targeting.....	113
Appendix 3 - <i>ccd4</i> alleles .....	114
Allele W <sup>1</sup> .....	114
Allele Y <sup>1</sup> .....	116
Allele Y <sup>2</sup> .....	118
Allele Y <sup>3</sup> .....	122
Bibliography .....	124