UNIVERSITAT AUTÒNOMA DE BARCELONA

Faculty of Bioscience

Department of Animal Biology, Plant Biology and Ecology

Identification and characterization of candidate genes for peach fruit shape

Thesis presented by Yu Zhang for the degree of Doctor in Plant Biology and Biotechnology of Universitat Autònoma De Barcelona (UAB).

This work was performed at the department of Plant Genetics of the Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Barcelona.

Thesis director   PhD candidate

María José Aranzana Civit   Yu Zhang

Barcelona, October 2018

To my parents,
To Xiaoqin, Huiyi, and Xinmi,
To Tingting.

Learning conceives ambition,
ambition gives rise to virtue,
To be diligent and frugal,
Pursue words and deeds in virtuous and committed.

Acknowledgements

this experience with working with you. Thank Xènia, a mart and lusty girl, for the experimental assistance and those helpful ideas that inspired me a lot to carry out the research. And also Xènia was involved in a part of my project work, which adds distinctive radiance to my thesis. Carlos and Jorge, thanks for your enthusiastic assistance in the lab, and also we are the Montserrat squad, you nice guys will be memorable in my life. Thank the fellow lab mates Covadonga Vara from Aurora lab, Naveen, Christian, Neus, Nuria, Pol, Maria, Pablo, for the explanation and translation of department or lab meeting, for the stimulating discussion of experiment, for the countless days we were working together, and for all the fun we have had in the last four years.

I acknowledge those people from IRTA: Ibo and Mourad helped me with project when I started my thesis research, and with the operation in apparatus, machines, and facilities. Àngel, Esteve, Fuensi, Sara, and Elena with abundant experience working in all stuffs in the lab, the greenhouse, the *in vitro*, have been supportive and aidant for my work. I would like to express my cordial appreciation to all of you.

The genuine thanks also goes to my kind and benign roommates and friends Rong, Yan, Wenjie, Changyong, Chunyi, Junpeng, Fangchang, Yulu, Xialei, Xiaoqing, Yaxing, Baoyi, Liubin, Xiaolang, Dailu, *etc.,* for your dedication and care when I had troubles during the ordinary life, for your valuable suggestions and concise comments on parts of my PhD work as well as thesis writing. I have learnt different and extraordinary skills from your various life experience, and your being by my side gave me positive motivation to be stronger and more excellent. You happened to the especial treasures for me in the last four years, which I will keep in mind forever.

Last but not the least, I acknowledge my family: my parents, elder sister, sister-in-law, younger brothers, and brothers-in-law who encouraged and accompanied me at life stages. You are the firmest backing for me when I left home thousands of miles away. Thank my family for understanding and supporting me spiritually, for holding with my childishness and ignorance consistently throughout my life in general. At this moment of my thesis completion, I long to share this cheer with you dearests. The sincerest gratitude and appreciation from the bottom of my heart is given to Tingting who has outstanding personal qualities and attractive characters. Within my work and thesis writing, thank you for accompanying me day and night unconditionally, for supporting my thoughts about research project and choices of future life, for conforming me when I presented self-accusation or sense of loss after failure.

Thank all the people who helped me, I am and will be eternally appreciated. Best wishes for all of you!

Index of contents

Index of figures

Index of tables

Abbreviations

A	Adenine
ABA	Abscisic acid
AFLP	Amplified Fragment Length Polymorphism
ASF	Agro Sélection Fruits
Asn	Asparagine
BAC	Bacterial Artificial Chromosome
BC	Backcross
BiNGO	Biological Networks Gene Ontology tool
BLAST	Basic Local Alignment Search Tool
BLOSUM	Blocks of Amino Acid Substitution Matrix
bp	base pair
C	Cytosine
cDNA	complementary DNA
CDS	Coding DNA Sequence
CE	Capillary Electrophoresis
CG	Candidate gene
cM	centiMorgan
CTAB	Cetyl Trimethylammonium Bromide
cv	cultivar
DAPI	4', 6-diamidino-2-phenylindole
DEGs	Differential Expression of Genes
DNA	Deoxyribonucleic acid
dNTP	deoxy Nucleotide Triphosphate
EST	Expressed Sequenced Tag
F1	First Fillial Generation

F2	Second Fillial Generation
FAOSTAT	Statistics division of the FAO (Food and Agriculture Organization)
G	Guanine
GDR	Genome Database Rosaceae
GFP	Green Fluorescent Protein
GO	Gene Onthology
GWAS	Genome-wide Association Study
HICF	High-information Content Fingerprinting
INDELs	Insertion/Deletion polymorphisms
IPTG	Isopropyl-β-D-1-tiogalactopiranoside
Kb	Kilo base
Kv	Kilo volt
Leu	Leucine
LG	Linkage Group
LOH	Loss of Heterozygosity
LRR	Leucine Rich Repeat
MAS	Marker Assisted Selection
Mb	Mega base
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
PK	Protein Kinase
PPV	Plum pox virus
QTLs	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length
RLKs	Receptor Like Kinases
RNA	Ribonucleic acid

RNAi	RNA interference
rpm	revolutions per minute
RT-PCR	Reverse-Transcript PCR
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
STR	Short Tandem Repeats
T	Thymine
TA	Titratable Acidity
TALEN	Transcription Activator-like Effector Nuclease
TE	Transposable Element
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TSS	Transcription Start Site
UTR	Untranslated Region
WGS	Whole Genome Shotgun
YAC	Yeast Artificial Chromosome
ZNF	Zinc-finger Nuclease

Summary

Peach fruit derives from the ripen ovary, developing into drupes whose shape may vary from round to flat, acquiring some intermediates shapes as the oblate. In peach, the flat phenotype is caused by a partially dominant allele in heterozygosis (*Ss*), fruit from homozygous trees (*SS*) abort a few weeks after fruit setting. Previous research has identified a SSR marker (UDP98-412) highly associated with the trait, found suitable for marker assisted selection (MAS). Later, an association analysis suggested the putative involvement of a leucine rich receptor-like kinase (LRR-RLK) in fruit shape determination. Specifically, a ~10Kb deletion affecting the promoter and part of the coding region of the gene co-segregated with the flat shape.

Here our goal was to clone and study in detail that gene (*Prupe.6G281100*) to elucidate its function in determining the shape of the fruit. This gene is orthologous to the *BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1* (*BAK1*). Protein BLAST alignment identified significant hits with genes involved in different biological processes. Best protein hit occurred with *AtRLP12*, which may functionally complement *CLAVATA2*, a key regulator that controls the stem cell population size. RT-PCR analysis revealed the absence of transcription of the partially deleted allele associated with the flat phenotype. The data support *Prupe.6G281100* as a candidate gene for flat shape in peach. The posterior screening of larger sample panel has identified three varieties escaping the association between genotype at *Prupe.6G281100* and the phenotype. The whole genome sequences of these three varieties revealed additional variability in other two LRR-RLKs genes clustering, together with *Prupe.6G281100,* in the *S* locus. Large fraction of this variability accumulated in one of these two genes (*Prupe.6G281500* located 21.3 Kb downstream *Prupe.6G281100).* This variability, in phase with the 10 Kb deletion, was observed in all flat varieties. The second gene with additional variability was *Prupe.6G281400* (15.3 Kb downstream *Prupe.6G28110)* with a 6.1 Kb sequences deleted in phase with the *S* allele and unique to the outlier varieties. According to gene function prediction, the two LRR-RLKs genes have best homology with *FLORAL ORGAN NUMBER1* (*FON1*) and *THICK TASSEL DWARF1* (*TD1*). These genes have been reported to be involved in mediating floral meristems and organs growth during inflorescence and flower development, including structural organization, shape alteration and size regulation in various species. We formulate here a three allele model hypothesis based on allele conformation at the three genes, which may be involved at different levels of dominance, in fruit shape determination.

In this thesis we have also analyzed a somatic mutant with round fruit 'UFO-4Mut' derived from the flat variety 'UFO-4'. Genotypic data revealed that this mutation occurred

in cells of the meristematic layer LII. We initiated the study of these two cultivars with the main objective of validating *Prupe.6G28000* role in fruit shape determination. However, whole genome analysis revealed a large region (6.5 Mb) in the mutant sample with loss of heterozygosity (LOH). We hypothesize that the LOH may have been produced by the repair of a double strand break (DSB) in the chromosome carrying the allele causing the flat phenotype with the homologous chromosome (carrying the allele for the round). Therefore, the allele producing flat shape has been deleted in the mutant and replaced by the one producing the round shape, which fully explains the phenotype reversion. In this thesis we also initiated protocols for two different techniques. One of them aims at the visualization of genome rearrangements in peach somatic cells by FISH. The other consists on the use of virus vectors to induce gene expression in peach. First steps in the setting of these protocols have been established and are discussed in this thesis manuscript.

Resum

El fruit del préssec deriva de l'ovari madur, que es converteix en drupa. La seva forma pot variar entre rodona i plana, adquirint algunes formes intermèdies com la forma aplanada. En el préssec, el fruit pla (anomenat paraguaià) és causat per un al·lel parcialment dominant que ha d'estar en heterocigosis (*Ss*), mentre que els fruits d'arbres homozigots (*SS*) avorten poques setmanes després del quallat del fruit. Investigacions anteriors han identificat un marcador SSR (UDP98-412) altament associat al caràcter i que és adequat per a la seva utilització en selecció assistida per marcadors (SAM). Posteriorment, una anàlisi d'associació va revelar la possible implicació d'una quinasa del tipus Leucine Reach Repeat Receptore Like Kinase (LRR-RLK) en la determinació de la forma del fruit. En concret, una deleció de ~ 10Kb que afecta el promotor i part de la regió codificant d'aquest gen co-segrega amb la forma plana.

Aquí, el nostre objectiu era clonar i estudiar en detall aquest gen (*Prupe.6G281100*) per dilucidar la seva funció en la determinació de la forma del fruit. Aquest gen és ortòleg al gen *BAK1* (BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1). Una anàlisi d'alineament (BLAST) entre proteïnes va identificar homologia significativa amb gens involucrats en diferents processos biològics. El millor alineament es va produir amb el gen AtRLP12, que pot complementar funcionalment *CLAVATA2*, un regulador clau que controla la mida de la població de cèl·lules del meristema. L'anàlisi per RT-PCR va revelar l'absència de transcripció de l'al·lel parcialment delecionat, associat amb el fenotip pla. Les dades donen suport a *Prupe.6G281100* com un gen candidat per a la forma plana en préssec. El genotipat posterior d'un panell més gran de varietats ha identificat tres varietats que escapen a l'associació entre el genotip a *Prupe.6G281100* i el fenotip. Les seqüències genòmiques completes d'aquestes tres varietats van revelar variabilitat addicional en altres dos de gens LRR-RLK situats en clúster, juntament amb *Prupe.6G281100*, al locus *S*. Gran fracció d'aquesta variabilitat estava acumulada en un d'aquests dos gens (*Prupe.6G281500* localitzat a 21.3 Kb de *Prupe.6G281100*). Aquesta variabilitat, en fase amb la deleció de 10 kb, es va observar en totes les varietats planes. El segon gen amb variabilitat addicional va ser *Prupe.6G281400* (a 15.3 Kb de *Prupe.6G281100*) amb una seqüència de 6.1 Kb eliminada en fase amb l'al·lel *S* i única per a les varietats amb fenotip discordant. D'acord amb la predicció funcional, els dos gens LRR-RLK tenen la millor homologia amb *NUMERO D'ORGANO FLORAL1* (*FON1*) i amb *DWARF1 DE GROSSA TASCA* (*TD1*). A la bibliografia es descriu que aquests gens estan involucrats en el desenvolupament dels meristemes florals i el creixement dels òrgans durant la inflorescència i el desenvolupament de les flors, inclosa l'organització estructural, l'alteració de la forma i la regulació de la mida en diverses espècies. A partir de les nostres dades, formulem aquí una hipòtesi de model de tres al·lels

basada en la conformació d'al·lels en els tres gens, que poden estar involucrats, amb diferents nivells de dominància, en la determinació de la forma del fruit.

En aquesta tesi també hem analitzat un mutant somàtic amb fruit rodó "UFO-4Mut" derivat de la varietat plana "UFO-4". Les dades genotípics van revelar que aquesta mutació es va produir en cèl·lules de la capa meristemàtica LII. Iniciem l'estudi d'aquests dos cultivars amb l'objectiu principal de validar el paper de *Prupe.6G28000* en la determinació de la forma del fruit. No obstant això, la resecuenciación dels dos genomes va revelar una gran regió (6.5Mb) amb pèrdua d'heterozigositat (LOH) en la mostra mutant. La nostra hipòtesi és que la LOH pot haver estat produïda per la reparació d'un trencament de doble cadena (DSB) ocorreguda en el cromosoma que porta l'al·lel que causa el fenotip pla, amb el cromosoma homòleg (que porta l'al·lel per al fruit rodó). Per tant, l'al·lel que produeix una forma plana s'ha eliminat en el mutant i reemplaçat pel que produeix la forma rodona, el que explica completament la reversió del fenotip. En aquesta tesi també vam iniciar el desenvolupament dels protocols per a dues tècniques diferents. Una d'elles té com a objectiu la visualització de reordenaments cromosòmics en cèl·lules somàtiques per FISH. L'altre consisteix en l'ús de vectors virals per induir l'expressió de gens en préssec. Hem establert els primers passos en l'establiment d'aquests protocols, els quals es discuteixen en aquest manuscrit de tesi.

Resumen

El fruto del melocotón deriva del ovario maduro, que se convierte en drupa. Su forma puede variar entre redonda y plana, adquiriendo algunas formas intermedias como la forma achatada. En el melocotón, el fruto plano (llamado paraguayo) es causado por un alelo parcialmente dominante que debe estar en heterocigosis (*Ss*), mientras que los frutos de árboles homocigotos (*SS*) abortan pocas semanas después del cuajado del fruto. Investigaciones anteriores han identificado un marcador SSR (UDP98-412) altamente asociado al carácter y que es adecuado para su utilización en selección asistida por marcadores (SAM). Posteriormente, un análisis de asociación reveló la posible implicación de una quinasa del tipo Leucine Reach Repeat Receptore Like Kinase (LRR-RLK) en la determinación de la forma del fruto. En concreto, una deleción de ~ 10Kb que afecta al promotor y parte de la región codificante de ese gen co-segrega con la forma plana.

Aquí, nuestro objetivo era clonar y estudiar en detalle ese gen (*Prupe.6G281100*) para dilucidar su función en la determinación de la forma del fruto. Este gen es ortólogo al gen *BAK1* (*BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1).* Un análisis de alineamiento (BLAST) entre proteínas identificó homología significativa con genes involucrados en diferentes procesos biológicos. El mejor alineamiento se produjo con el gen AtRLP12, que puede complementar funcionalmente *CLAVATA2*, un regulador clave que controla el tamaño de la población de células del meristemo. El análisis por RT-PCR reveló la ausencia de transcripción del alelo parcialmente delecionado, asociado con el fenotipo plano. Los datos respaldan a *Prupe.6G281100* como un gen candidato para la forma plana en melocotón. El genotipo posterior de un panel mayor de variedades ha identificado tres variedades que escapan a la asociación entre el genotipo en *Prupe.6G281100* y el fenotipo. Las secuencias genómicas completas de estas tres variedades revelaron variabilidad adicional en otros dos de genes LRR-RLK situados en cluster, junto con *Prupe.6G281100*, en el locus *S*. Gran fracción de esta variabilidad estaba acumulada en uno de estos dos genes (*Prupe.6G281500* localizado a 21.3 Kb de *Prupe.6G281100*). Esta variabilidad, en fase con la deleción de 10 kb, se observó en todas las variedades planas. El segundo gen con variabilidad adicional fue *Prupe.6G281400* (a 15.3 Kb de *Prupe.6G28110*) con una secuencia de 6.1 Kb eliminada en fase con el alelo S y única para las variedades con fenotipo discordante. De acuerdo con la predicción funcional, los dos genes LRR-RLK tienen la mejor homología con *NUMERO DE ORGANO FLORAL1* (*FON1*) y con *DWARF1 DE GRUESA TASCA (TD1).* En la bibliografía se describe que estos genes están involucrados en el desarrollo de los meristemos florales y el crecimiento de los órganos durante la inflorescencia y el desarrollo de las flores, incluida la organización estructural, la alteración de la forma y la regulación del tamaño en diversas especies. A partir de nuestros datos, formulamos aquí una hipótesis de modelo de tres alelos basada en la conformación de alelos en los tres

genes, que pueden estar involucrados, con diferentes niveles de dominancia, en la determinación de la forma del fruto.

En esta tesis también hemos analizado un mutante somático con fruto redondo "UFO-4Mut" derivado de la variedad plana "UFO-4". Los datos genotípicos revelaron que esta mutación se produjo en células de la capa meristemática LII. Iniciamos el estudio de estos dos cultivares con el objetivo principal de validar el papel de Prupe.6G28000 en la determinación de la forma del fruto. Sin embargo, la resecuenciación de los dos genomas reveló una gran región (6.5Mb) con pérdida de heterocigosidad (LOH) en la muestra mutante. Nuestra hipótesis es que la LOH puede haber sido producida por la reparación de una rotura de doble cadena (DSB) ocurrida en el cromosoma que porta el alelo que causa el fenotipo plano, con el cromosoma homólogo (que lleva el alelo para el fruto redondo). Por lo tanto, el alelo que produce una forma plana se ha eliminado en el mutante y reemplazado por el que produce la forma redonda, lo que explica completamente la reversión del fenotipo. En esta tesis también iniciamos la puesta a punto de protocolos para dos técnicas diferentes. Una de ellas tiene como objetivo la visualización de reordenamientos del genoma en cromosomas de células somáticas por FISH. El otro consiste en el uso de vectores virales para inducir la expresión de genes en melocotón. Hemos establecido los primeros pasos de estos protocolos se han establecido, los cuales se discuten en este manuscrito de tesis.

General introduction

1 Peach origin and germplasm

Peach (*Prunus persica* [L.] Batsch) is a perennial deciduous tree species. It belongs to the *Prunus* genus within the *Amygdaloideae* subfamily in the Rosaceae family (Potter *et al.*, 2007). Peach encompasses an ample number of commercial cultivars, grown in the range of 30° - 45° N and 30° - 45° S latitudes (Rehder, 1940). Peaches were native to Northwest China between Tarim Basin and the north of the Kunlun Mountains, where they have been cultivated for more than 4000 years (Hedrick *et al.*, 1917; Wang, 1985; Faust & Timon, 1995), since rainwater was abundant in that area thousands of years ago. In days of yore, farmers domesticated peach fruit along with pears and apricots long before the Zhou Dynasty period (1046 - 256 BC) in China (Li, 1983). The earliest literary inscription about peach cultivation refers to Shijing (Classic of Poetry) written around 1000 - 500 BC, acclaiming the beauty of peach blossoms and the fragrance of peach fruit (Li, 1984).

The large output of China's peach germplasm resources can be traced back to the period of Emperor Wu of the Han Dynasty (141 - 87 BC). At that time, as the diplomat and imperial envoy Zhang Qian visited the Western Region from Chang'an (Xi'an nowadays), peaches were spread to Persia and to the West Asian countries while he was opening the Silk Road; later the Romans brought them from Persia to the European Mediterranean countries (Wang & Zhuang, 2001; Xie *et al.*, 2007; Byrne *et al.*, 2012). With the discovery of the New World, peaches were introduced to Mexico and South America by the Portuguese and the Spanish around the 16th century, finally peach germplasm resources spread all over the world (Layne & Bassi, 2008; Chin *et al.*, 2014).

Chinese peach germplasm resources have provided important support for the development of the world peach breeding. In the modern times with some qualified fruit characteristics, the emergence of 'Shuimitao' ('honey peach' from China) has become the one of the pillars of modern peach breeding worldwide (Wang & Zhuang, 2001). Peach 'Chinese Cling' ('Shanghaishuimi'; Figure I.1) was introduced in Georgia (USA) by mid-19th century and intensively used in breeding programs, contributing to the current composition of the world peach varieties, as most of the cultivars in the world are directly or indirectly derived from it (Janick & Moore, 1996). From 'Chinese Cling' derived 'Belle of Georgia', 'Belle', 'Elberta' among other highly popular varieties. For the following half century 'Elberta' and 'J. H. Hale' (supposedly a seedling of 'Elberta') become the most important cultivars in the US (Hedrick, 1950; Roach, 1985). By using them as recurrent parental lines in breeding programs, the US selected and cultivated more than 500 peach cultivars, including yellow peaches 'Babygold', 'Redhaven', and nectarines 'Sunfre', 'Flordared', which have been widely cultivated throughout the world (Scorza *et al.*, 1985; Myers *et al.*, 1989).

Similar situation occurred in other parts of the Worlds. For example, with the introduction by 1875 of 'Chinese Cling', 'Tianjinshuimi' as well as other large fruit-shaped varieties, Japan bred 'Tuyongshuimi' ('Doyou'), 'Liheshuimi' and 'Gangshanbai' ('Hakuho'), and soon 'Gangshanbai' was used to obtain 'Baifeng' and 'Dajiubao' ('Okubo'), which are the bases of the main cultivars of modern peaches in that country (Wang & Zhuang, 2001; Li, 2013), fitting Japanese consumers' taste for low-acid varieties. In Italy high-scale peach breeding programs, based only in introductions from US (particularly 'J.H.Hale'), were initiated in the late 1920s by Armando Morettini of the University of Florence (Pirovano, 1953). However Italian germplasm resources are reach, encompassing local cultivars that have been grown at certain extent such as the white peaches 'Alibance', 'Alirosada', 'Maria Bianca', 'Rubia', 'Rubisco'. Similarly, Spanish peach industry has relied for years in the use of introduced cultivars, usually poorly adapted to Spanish environment (Llácer *et al.*, 2009). During the last decades of 20th century large-scale breeding programs were set up in Spain from US introductions; these programs constitute now a referent in peach breeding worldwide (Iglesias, 2018). Local Spanish cultivars, are currently preserved in germplasm collections. Among some poplar cultivars we might cite the canning peaches 'Sandanell', 'Amarillo Tardío de Calanda', 'Jerónimo', 'San Lorenzo' and 'Campiel', which are maintained in germplasm collections as the one in the Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA).
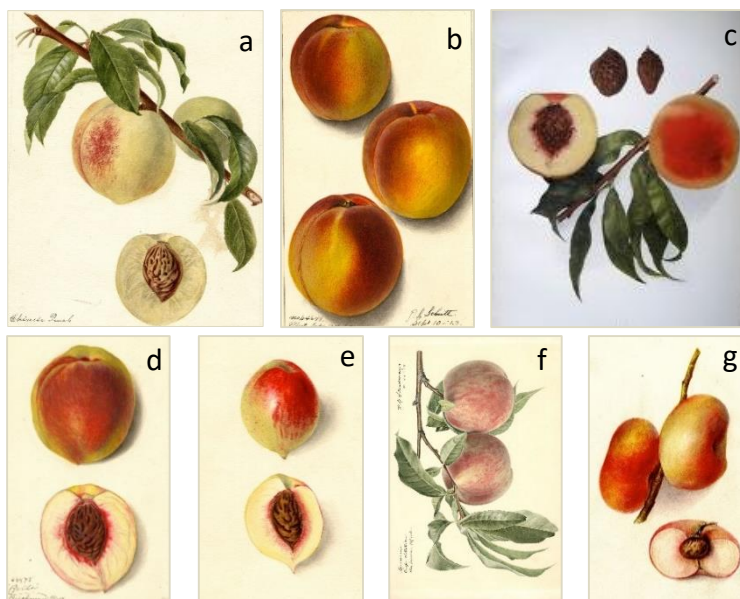
Figure I.1 Ancient and mordern typic peach germplasm resources. **a)** 'Chinese cling', one of the most important breeding stock varieties. (Image from Deborah Passmore, 1893.) **b)** 'Elberta', germplasm collection PI673789 retained in the USDA. (Image from Ethel Schutt, 1913.) **c)** 'J.H. Hale', supposedly a seedling of 'Elberta' **d)** 'Belle of Georgia', USDA PI 673710 visually one of the most pleasing of peaches. (Image from Amanda Newton, 1913) **e)** 'Honey', the origin has never been definitively

ascertained. (Image from Elsie Lower, 1910.) **f)** CARMAN , One of the North China configuration (Imagefrom Royal Steadman, 1917.) **g)** 'Peetao', USDA PI 673610, ancestor to the present day doughnut peaches.

2 Peach production and economic value

Peach fruit, developed from the enlargement of the ovary, provide humans and animals with source of food. Peach fruit nutritional value include vitamins, carotenoid, phenolic compounds and dietary fiber, with its health benefits associated (Ebihara *et al.*, 1979; Wills *et al.*, 1983). Thanks to such benefits, peach has been an economically important fruit, which took up the tenth fruit species status in global production quantity, with 24.98 million tons harvested covering 1.64 million hectares in 2016. European countries and regions produced 4.37 million metric tons of peaches within a harvest area 0.26 million ha, only surpassed by apple and orange. In 2016 China the top peach producing country yielded 14.47 million tons, occupying a portion of 58% of the world's total peach production. Spain produced 1.53 million tons, holding in the second followed by Italy with 1.43 million tons. The posterior ranking is US with 0.93 million tons, Iran was the fifth largest peach producing country producing 0.86 million tons (Figure I.2) (FAOSTAT, 2016).
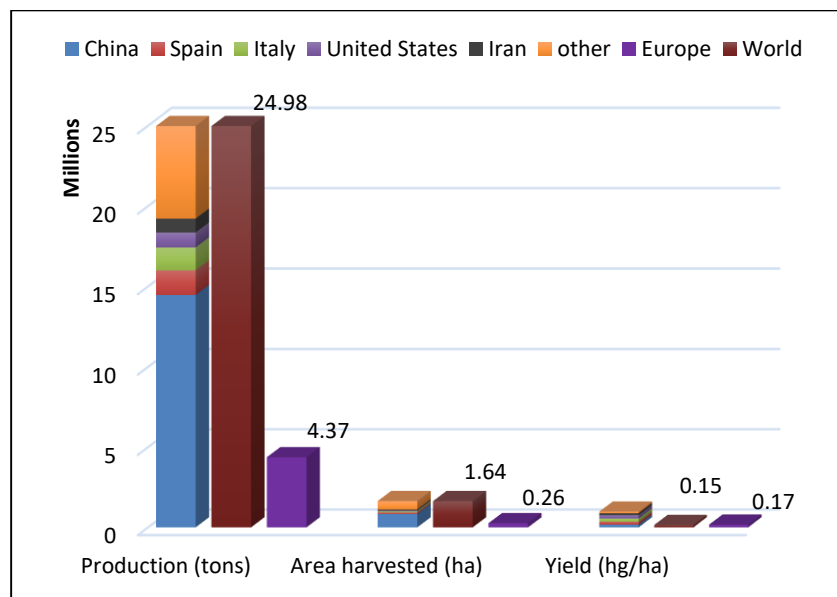
Figure I.2 Global peach production, harvested area and yield in 2016. Comparison between World, Europe and the first five countries are identified by diacritical colors on the histogram. (data from FAOSTAT*)*

During 1961 – 2016, the global peach production showed a rising trend, although with some periods of slight decay (Figure I.3). Similar behavior was observed in the Spanish

production system (FAOSTAT, 2016). This increase in production and harvested area in Spain has been accompanied by an impressive varietal renewal, with peaches with better quality and better adapted to Spanish agro-climatic conditions (Figure I.4). This has been possible thanks to the establishment of more than ten breeding companies in Spain in the last three decades; nowadays Spain is world leader in variety renewal (Iglesias, 2018).
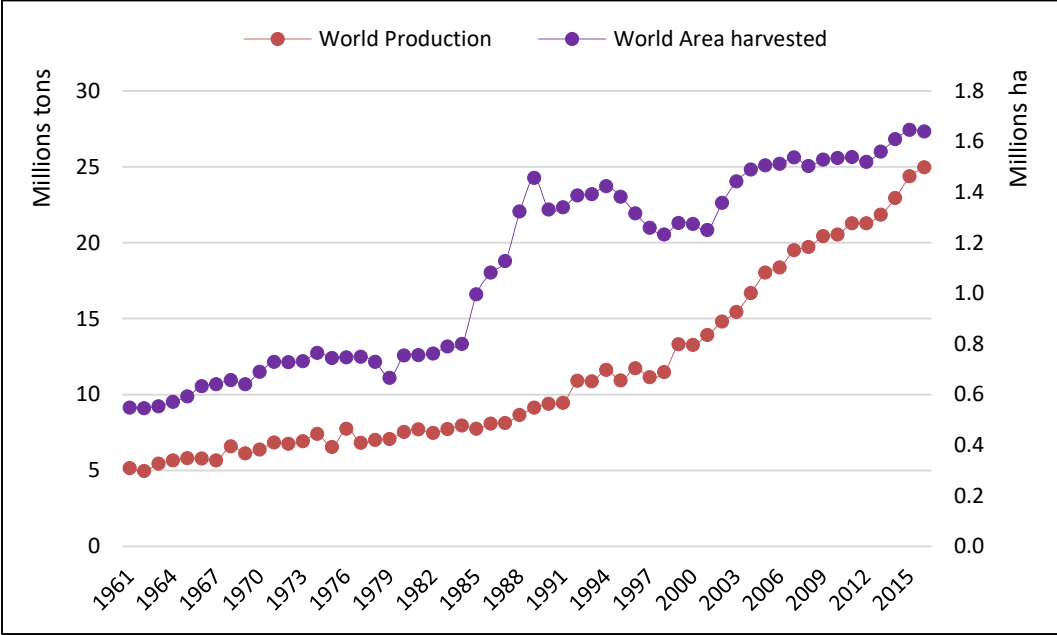
Figure I.3 Global peach production and cultivated area during 1961 – 2016. (data from FAOSTAT)
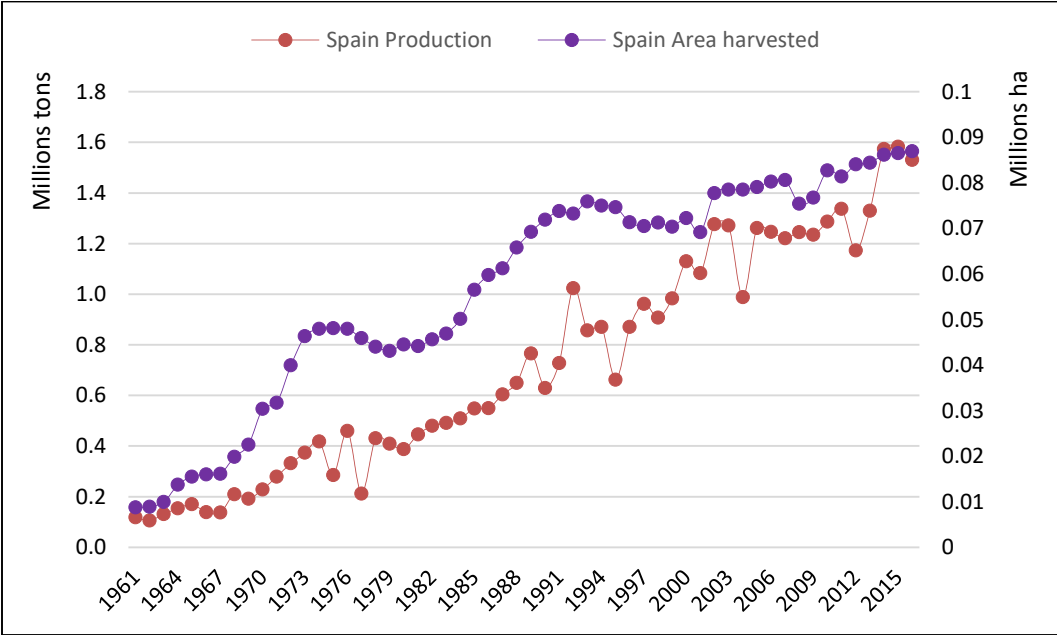
Figure I.4 Spain peach production and cultivated area during 1961 – 2016. (data from FAOSTAT)

3 Peach genomics

3.1 Linkage map construction

Peach is a diploid stone fruit species with an average 2-3 years of intergenerational period that is relatively short compared to most fruit tree species. Peach, with a small genome size (~227 Mb) and chromosomal base number (2n = 16), is considered model species for genomic studies in the Rosaceae family (Shulaev *et al.*, 2008). Therefore, scientific efforts have provided numerous and useful tools for peach and other *Rosaceae* genetic studies. Among them we may highlight the genetic maps, i.e. the relative location of molecular markers or genes on a chromosome based on recombination frequencies in a pedigree related progeny.

In the last twenty years, great progress has been made in linkage maps construction, which have allowed a big progress in the research of gene localization, comparative genomics, germplasm resource evaluation and molecular assisted selection (MAS) breeding. In 1994, Chaparro and collaborators published the first peach linkage map constructed with 96 individuals of 'NC174RL' × 'Pillar' F2 progeny (Chaparro *et al.*, 1994). This map was constructed with one isozyme, four morphological traits and 83 RAPD markers, which were assigned to 15 linkage groups, covering approximately 396 cM with an average density of 4.8 cM per marker (Table I.1). Soon afterwards, Joobeur *et al.* published a linkage map using the interspecific F2 population from crossing almond ('Texas') × ('Earlygold') using 235 RFLPs and 11 isozymes (Table I.2). The T×E mapped covered a total distance of 491 cM with an average 2.0 cM/marker(Joobeur *et al.*, 1998). This map was considered soon the reference map for *Prunus* species and was subsequently improved thanks to the development of new marker types like RAPD, AFLP and SSR (Joobeur *et al.*, 2000; Aranzana *et al.*, 2003; Howad *et al.*, 2005; Picañol *et al.*, 2013). The recently developed Single Nucleotide Polymorphism (SNP) markers have become more practical and functional in linkage map construction due to low mutation rate and stability from generation to generation across the genome (Batley & Edwards, 2009). By using next generation sequencing platforms, the International Peach SNP Consortium (IPSC) resequenced 56 peach varieties and identified one million SNPs, generating the 9K peach SNP array v1 used for peach and related stone fruit species research (Verde *et al.*, 2012). These SNP array was used to genotyped the TxE as well as other populations derived from it, allowing obtaining highly saturated maps (Donoso *et al.*, 2015; Donoso *et al.*, 2016). Similarly, Martínez-García *et al.* used 1536 SNPs in a peach x peach and in a peach x (peach x almond) hybrid population to develop a consensus map assigned to eight linkage groups covering 454 cM with a density of 0.81 cM/marker site (Martínez-García *et al.*, 2013). Recently, the IPSC has

developed a new array containing 9,000 additional SNPs (18K SNP array) and has started to be supplied by Illumina company. In spite that more efficient tools for genomics were/are under development over the years, SNP marker is considered the most ideal method for peach saturated linkage map construction at present.

Table I.1 Peach linkage maps constructed with intraspecific cross.

Mapping population		Marker types		Linkage groups	Genetic distance (cM)		References
Parents	Population type	Markers	Total		Total	Average	
'NC174RL' x 'Pillar'	F2 (160)	1 Isozyme, 4 Morphological, 83 RAPDs	88	15	396	4.8	(Chaparro *et al.*, 1994)
'New Jersey Pillar' × 'KV77119'	F2 (71)	7 Morphological, 46 RFLPs, 12 RAPDs	65	8	332	8	(Rajapakse *et al.*, 1995)
'Ferjalou Jalousia' × 'Fantasia'	F2 (63)	1 Isozyme, 4 Morphological, 6 IMAs, 82 RAPDs, 47 RFLPs, 109 AFLPs	249	11	712	4.5	(Dirlewanger *et al.*, 1998)
'Lovell' × 'Nemared'	F2 (55)	AFLPs	153	15	1297	9.1	(Lu *et al.*, 1998)
'Akame' × 'Juseitou'	F2 (126)	9 Morphological, 31 RAPDs, 1 RFLPs, 35 AFLPs, 11 SSRs, 5I SSRs	92	9	1020	12	(Yamamoto *et al.*, 2001)
'Akame' × 'Juseitou'	F2 (126)	9 Morphological, 24 RAPDs, 34 AFLPs, 94 SSRs, 3 ISSRs, 14 STSs	178	8	571	3.2	(Yamamoto *et al.*, 2005)
'Ferjalou Jalousia' × 'Fantasia'	F2 (207)	1 Isozyme, 6 Morphological, 37 RFLPs, 61 AFLPs, 82 SSRs	187	7	621.2	3.3	(Dirlewanger *et al.*, 2006)
'Guardian® 3-17-7' × 'Nemaguard'	F2 (100)	140 AFLPs, 18 SSRs	158	11	734	4.7	(Blenda *et al.*, 2007)
'Dr. Davis' × 'Georgia Belle'	F2 (152)	3 Morphological, 24 CGs, 79 SSRs, 40 RAFs, 23 SRAPs, 14 IMAs, 28 others	211	8	818.2	4	(Ogundiwin *et al.*, 2009)
'Venus' × 'Big Top'	F1 (75)	5 SSRs, 99 SNPs ('Venus')	104	9	129.9	2.49	(Zeballos *et al.*, 2016)
		1 SSRs, 122 SNPs ('Big Top')	123	10	464.3	3.8	
'YM' × 'HJML'	F1 (103)	SNPs	1,310	8	454.2	0.347	(Guo *et al.*, 2018)

Table I.2 Peach linkage maps constructed with interspecific cross.

Mapping population		Marker types		Linkage groups	Genetic distance (cM)		References
Parents	Population type	Markers	Total		Total	Average	
Peach '54P455' × Almond 'Padre'	F2 (64)	6 Isozymes, 1 Morphological, 101 RFLPs	108	9	800	7.5	(Foolad *et al.*, 1995)
Almond 'Texas' × Peach 'Earlygold'	F2 (111)	11 Isozymes, 235 RFLPs	246	8	491	2	(Joobeur *et al.*, 1998)
'IF7310828' × ('IF7310828' × '*P. ferganensis*')	BC1 (76)	RFLP, RAPD	36	8	257	7.6	(Quarta *et al.*, 1998)
'IF7310828' × ('IF7310828' × '*P. ferganensis*')	BC1 (70)	2 Morphological, 16 RAPDs, 74 RFLPs, 17 SSRs	109	10	521	4.8	(Dettori *et al.*, 2001)
Almond 'Garfi' × Peach 'Nemared'	F2 (113)	5 Isozymes, 46 RFLPs	51	7	-	-	(Jáuregui *et al.*, 2001)
Almond 'Padre' × Peach '54P455'	F2 (64)	8 Isozymes, 4 Morphological, 1 RAPD, 143 RFLPs, 4 SSRs, 1 CAP	161	8	1144	6.8	(Bliss *et al.*, 2002)
Almond 'Texas' × Peach 'Earlygold'	F2 (111)	11 Isozymes, 235 RFLPs, 96 SSRs	352	8	522	5.4	(Aranzana *et al.*, 2003)
Almond 'Texas' × Peach 'Earlygold'	F2 (111)	11 Isozymes, 361 RFLPs, 185 SSRs, 5 ESTs	562	8	519	0.92	(Dirlewanger *et al.*, 2004a)
Almond 'Texas' × Peach 'Earlygold'	F2 (111)	11 Isozymes, 361 RFLPs, 449 SSRs, 5 ESTs	826	8	524	0.63	(Howad *et al.*, 2005)
'Bailey' × ('Honggengansutao' × 'Bailey')	BC1 (190)	1 Morphological, 30 SSRs, 102 SRAPs, 5 RGA-STSs	138	8	616	4.9	(Cao *et al.*, 2011)
Pop-DF ('Dr. Davis' × 'F8,1-42'), Pop-DG ('Dr. Davis' × 'Georgia Belle')	Pop-DF (69F1) Pop-DG (55F1)	SNPs	588	8	454.8	0.81	(Martínez-García *et al.*, 2013)
Almond 'Texas' × peach 'Earlygold'	F2 (111)	114 SSRs, 1834 SNPs (T × E)	1,948	8	472.1	-	(Donoso *et al.*, 2015; Donoso *et al.*, 2016)
'Earlygold' × ('Texas' × 'Earlygold')	BC1 (190)	113 SSRs, 1919 SNPs (T1E)	2,032	8	370.1	-	
		40 SSRs, 1050 SNPs (E)	1,091		520.4	-	

3.2 Mapping major genes and QTLs for agronomic traits

The co-localization of molecular markers and agronomic traits in linkage maps has become an effective method for identifying genomic regions encompassing genes for agronomic traits. At present, plentiful of these traits have been mapped, with different levels of accuracy, on the T×E *Prunus* reference map. This has allowed, in some cases, the development of markers for marker-assisted selection (MAS) in breeding. Among these mapped major genes, a significant number determine fruit quality, fruit type, plant/organ morphology and resistance to diseases and pathogens (Table I.3).

Most of these traits show simple Mendelian inheritance patterns. Some important peach fruit quality traits, in terms of appearance and taste, have been mapped as major genes. Among them we may cite skin pubescence (peach/nectarine, *G/g*), fruit flesh color (relow/white, *Y/y*), red flesh color (red/non-red flesh, *BF/bf* and *BF2/bf2*), red flesh color around stone (red/non-red, *Cs/cs*), flesh adhesion (clingstone/freestone, *F/f*), fruit type (almond/peach shape, Alf/alf), juicy flesh, (juicy/non-juicy, *Jui/jui*) and flat shape (flat/non-flat, *S/s*) (Yamamoto *et al.*, 2001; Bliss *et al.*, 2002; Arús *et al.*, 2012). Some taste traits have also been mapped as major genes, for example acidity (low-acid/ acid, *D/d*) and kernel taste (bitter/sweet, *SK/Sk*) (Bliss *et al.*, 2002; Boudehri *et al.*, 2009). However, other fruit quality traits are determined by various genes explaining each of them the observed phenotype in a major or minor measure. These genes have been mapped as QTLs in different linkage groups; among these traits we may highlight soluble-solid content, soluble-sugar content, sugar-acid composition (Quarta *et al.*, 2000; Etienne *et al.*, 2002). Among plant morphology characters we highlight here leaf color (red/green, *Gr/gr*) (Chaparro *et al.*, 1994; Yamamoto *et al.*, 2001), leaf shape (narrow/wide, *Nl/nl*) (Yamamoto *et al.*, 2001), flower color (pink/pale pink, *Fc/fc*) (Yamamoto *et al.*, 2001), double flower (single/double, *Dl/dl*) (Yamamoto *et al.*, 2001), leaf gland (reniform/globose/eglandular, *EE/Ee/ee*) (Quarta *et al.*, 2000; Dettori *et al.*, 2001), polycarpel (single/fused, *Pcp/pcp*) (Bliss *et al.*, 2002), pollen sterility (fertile/sterile, *Ps/ps*) (Dirlewanger *et al.*, 2006), plant height (normal/dwarf, *Dw/dw*) (Yamamoto *et al.*, 2001; Cantín *et al.*, 2018), and flower morphology (showy/non-showy, *Sh/sh*) (Fan *et al.*, 2010) (Table I.3).

Among pest and disease resistances traits, that to root-knot nematode is one of the most studied in peach. So far there are 4 root-knot nematode species described affecting peach: Meloidogyne incognita, M. javanica, M. arenaria and the recently reported M. floridensis. Genetic studies have identified genes providing resistance to one or more of these species. The gene *Mi* confers resistance to *M. incognita*, *Mj* to *M. javanica*, R_{MiaNem} to *M. incognita* and *M. arenaria*, and *Mf* to *M. floridensis*., Abbott *et al*

found that *Mi* and *Mij* were both mapped onto a single linkage group 1 (G1) (Abbott *et al.*, 1998). Likewise, Lu *et al.* reported a two-gene model that resistance to *M. incognita* and resistance to *M. javanica* were controlled by two dominant genes (*Mi* or *Mij*; and *Mj* or *Mij*, respectively) (Lu *et al.*, 1998; Lu *et al.*, 1999). By using an intraspecific F2 population crossed between 'Akame' and 'Juseitou', two new root-knot nematode resistance genes (*Mia* and *Mja*) were mapped on a different linkage group (G2) tightly linked together. In addition, another major gene R_{MiaNem} also localized in G2 with SSRs was identified to control resistance to *M. incognita* and *M. arenaria* (Claverie *et al.*, 2004a; Dirlewanger *et al.*, 2004b)*.* Recently, a newly peach pest disease, the peach root-knot nematode *M. floridensis* (Mf), was discovered to be not controlled by resistance genes in 'Nemared, 'Nemaguard' and 'Okinawa' cultivars (Maquilan *et al.*, 2018). Another disease highly studied in peach is powdery mildew, caused by *Sphaerotheca pannosa var. persicae*. Foulonge *et al.* identified thirteen quantitative trait loci (QTLs) conferring resistance in a major or minor way in three related F1, F2, and BC2 populations derived from crossing a susceptible peach pollen parent with a resistant *P. davidiana* seed parent (Foulongne *et al.*, 2003). A major gene (*Vr2*) was later identified by Pascal *et al.*, localized on G6 with the favorable allele coming from the peach cultivar 'Pamirskij 5'; the resistance occurred to be associated with the leaf color trait, facilitating its selection (Pascal *et al.*, 2010). Several QTLs have also been identified for *Plum pox virus* (PPV) disease resistance. The clone P1908 was used to study this complex, polygenic trait. Altogether, six QTLs in G1, G2, G4, G5, G6 and G7 were reported to present genes as a cluster of QTLs that confer resistance to PPV (Decroocq *et al.*, 2005; Marandel *et al.*, 2009; Rubio *et al.*, 2010).

In conclusion, a large fraction of peach tree and fruit traits are controlled by major or single genes, which facilitates the identification of molecular markers associated to those traits and posterior transfer to breeding programs for early selection of seedlings, what is known as marker assisted selection (MAS).

Table I.3 Peach major genes for agronomic traits.

Traits and major genes	Mapping populations	Markers types	Linkage groups	References
Plant morphology traits				
Double flower (single/double, *Dl/dl* or *Di/di*)	'NC174RL' x 'Pillar'	RAPD	G2 (*Dl/dl*)	(Chaparro *et al.*, 1994)
	'New Jersery Pillar' × 'KV77119'	SSR	G1 (*Dl/dl*)	(Sosinski *et al.*, 2000)
	'Weeping Flower Peach' × 'Pamirskij 5'	SSR	G6 (*Di2/di2*)	(Pascal *et al.*, 2017)
Flower color (pale pink/pink, *Fc/fc*)	'Akame' × 'Juseitou'	AFLP	G4	(Yamamoto *et al.*, 2001)
Flower morphology (showy/non-showy, *Sh/sh*)	'Contender' × 'Fla.92-2C'	AFLP, SSR	G8	(Fan *et al.*, 2010)
	'NC174R' x 'Pillar'	RAPD	G5	(Chaparro *et al.*, 1994)
Leaf color (red/green, *Gr/gr*)	'Akame' × 'Juseitou'	SSR	G3	(Yamamoto *et al.*, 2001)
	'Pamirskij 5' × 'Rubira®'	SSR	G6	(Lambert & Pascal, 2011)
Leaf gland (reniform/globose/eglandular, *EE/Ee/ee*)	'IF7310828' × '*P. ferganensis*'	RFLP	G7	(Quarta *et al.*, 2000)
	'IF7310828' × ('IF7310828' × '*P. ferganensis*')	RFLP	G7	(Dettori *et al.*, 2001)
Leaf shape (narrow/wide, *Nl/nl*)	'Akame' × 'Juseitou'	AFLP	G6	(Yamamoto *et al.*, 2001)
Male sterility (fertile/sterile, *Ps/ps*)	'Ferjalou Jalousia' × 'Fantasia'	AFLP	G8	(Dirlewanger *et al.*, 2006)
	'Texas' × 'Earlygold'	SSR, SNP	G2-G6	(Donoso *et al.*, 2015)
Plant height (normal/dwarf, *Dw/dw*)	'Akame' × 'Juseitou'	AFLP	G3	(Yamamoto *et al.*, 2001)
	'KV040127', 'KV011001'	SNP	G6	(Hollender *et al.*, 2016)

	'Nectavantop'	SSR, SNP	G6	(Cantín *et al.*, 2018)
Polycarpel (single/fused, *Pcp/pcp*)	'Padre' × '54P455'	Isozyme	G3	(Bliss *et al.*, 2002)
Resistance traits				
Root-knot nematode resistance (*Mij, Mi*)	'Lovell' × 'Nemared'	STS	G1	(Lu *et al.*, 1999)
Root-knot nematode resistance (*Mi/mi, Mj/mj*)	'Akame' × 'Juseitou'	ISSR	G2	(Yamamoto *et al.*, 2001)
Root-knot nematode resistance (*Mi/mi*)	'Harrow Blood' × 'Okinawa'	CAP	G2	(Gillen & Bliss, 2005)
Root-knot nematode resistance (*Mia, Mja*)	'Akame' × 'Juseitou'	SSR, STS	G2	(Yamamoto *et al.*, 2001)
Root-knot nematode resistance (*M$_{MiaNem}$*)	'Nemared' × 'Shalil'	SSR	G2	(Claverie *et al.*, 2004a)
Root-knot nematode resistance (*MF/mf*)	'*P. persica*' × '*P. kansuensis* Rehder'	SSR	G2	(Maquilan *et al.*, 2018)
Powdery mildew resistance (*Vr2/vr2*)	'IF7310828' × '*P. ferganensis*'	RFLP	G7	(Quarta *et al.*, 2000)
	'Rubira' × 'Pamirskij 5'	-	G6	(Pascal *et al.*, 2010)
	'Weeping Flower Peach' × 'Pamirskij 5'	SSR	G8	(Pascal *et al.*, 2017)
Plum pox virus resistance	'Summergrand' × 'P1908'	SSR	G1-G2-G4-G6-G7	(Decroocq *et al.*, 2005)
	A panel of accessions	SNP	G2-G3	(Cirilli *et al.*, 2017)
Fruit quality traits				
Aborting fruit (non-aborting/aborting, *Af/af*)	'Ferjalou Jalousia' × 'Fantasia'	SSR	G6	(Dirlewanger *et al.*, 2006)
Flesh adhesion (clingstone/freestone, *F/f*)	'Baily' × 'Suncrest'	RFLP	G2	(Abbott *et al.*, 1998)
	'IF7310828' × ('IF7310828' × '*P. ferganensis*')	RFLP	G4	(Dettori *et al.*, 2001)
	'IF7310828' × '*P. ferganensis*'	RFLP	G4	(Quarta *et al.*, 2000)
Flesh color (white/yellow, *Y/y*)	'Padre' × '54P455'	RFLP	G1	(Bliss *et al.*, 2002)
	'Redhaven', 'White Redhaven'	SSR, SCAR, SNP	G1	(Adami *et al.*, 2013)
Flesh color around the stone (red/white, *Cs/cs*)	'Akame' × 'Juseitou'	RAPD	G3	(Yamamoto *et al.*, 2001)

Trait	Cross / Population	Marker	Group	Reference
Flesh non-acid (low-acid/normally-acid, *D/d*)	'Ferjalou Jalousia' × 'Fantasia'	RAPD	G5	(Dirlewanger *et al.*, 1998)
	'Ferjalou Jalousia' × 'Fantasia'	SSR	G5	(Boudehri *et al.*, 2009)
	A panel of accessions	SSR, SNP	G5	(Eduardo *et al.*, 2014)
Flat shape (flat/round, *S/s*)	'Ferjalou Jalousia' × 'Fantasia'	SSR	G6	(Dirlewanger *et al.*, 2006)
	'UFO-3' × 'Sweet cap'	SSR	G6	(Picañol *et al.*, 2013)
Fruit maturity day (early/intermediate/late, *MD*)	'Summergrand' × 'P1908'	AFLP, SSR	G4	(Quilot *et al.*, 2004)
	'Contender' × 'Ambra', 'N.J. Weeping' × 'Bounty'	SSR	G4	(Pirona *et al.*, 2013)
Kernel taste (bitter/sweet, *SK/Sk*)	'Padre' × '54P455'	RFLP	G5	(Bliss *et al.*, 2002)
Skin color (red/green, *SC/Sc*)	'Akame' × 'Juseitou'	SSR	G3	(Yamamoto *et al.*, 2001)
	'IF7310828' × '*P. ferganensis*'	SSR	G2-G6	(Verde *et al.*, 2002)
Skin hairiness (nectarine/peach, *G/g*)	'Ferjalou Jalousia' × 'Fantasia'	AFLP	G5	(Dirlewanger *et al.*, 1998)

3.3 Physical maps

With the development of DNA sequencing techniques, physical maps become an important tool for genomic studies. Unlike linkage maps, which display genetic recombination rates among loci, physical maps are the alignment of DNA sequences, and therefore indicate the distance in base pairs between markers or genes. Integrated physical and genetic maps have provided essential support for gene fine mapping, map-based cloning, comparative genomics, high-throughput EST/SNP mapping, as well as for whole genome sequencing (Zhang & Wing, 1997; Green, 2001; Zhang & Wu, 2001).

The first peach physical map was constructed and released in 2006 using restriction-based fingerprinting method (Coulson *et al.*, 1986), and was established by employing strategies development for *Arabidopsis thaliana* (Marra *et al.*, 1999) and *Drosophila melanogaster* (Hoskins *et al.*, 2000). In peach, two large-insert genomic libraries based on BAC (bacterial artificial chromosome) vectors allowed to construct physical maps on varieties/populations with various phenotypes. Georgi *et al.* (2002) used DNA from leaves of the peach rootstock 'Nemared' to construct a BAC library containing 44,160 BAC clones with an average insert length of 70 kb and covering eight-fold of the haploid peach genome. Soon later, the 'Nemared' library was complemented with a second one constructed with the haploid peach strain 'Plov2-1 N' derived from the rootstock 'Lovell' (Zhebentyayeva *et al.*, 2006). The survey of the 'Nemared' library with AFLP and RFLP markers, most already mapped, allowed assigning mapped molecular markers to physical map contigs. In addition, the sequence of one of the BAC clones of close to 50,000 bp was used to develop new SSR markers. On the other hand, successive screenings of 'Nemared' BAC clones with additional molecular markers allowed to construct a candidate gene database and a transcript map for peach (Horn *et al.*, 2005). In a more exhaustive analysis, Zhenbentyayeva *et al.* development a high density BAC-based physical map for peach by fingerprinting the 'Nemared' and 'Lovell' libraries (Zhebentyayeva *et al.*, 2006). Shortly after, Zhebentyayeva *et al.* published the second peach physical map with higher number of BAC clones and higher accuracy of contig assembly (Zhebentyayeva *et al.*, 2008). For this they used high-information content fingerprinting (HICF), consisting on digesting the BAC clones with five restriction enzymes (*Bam*HI, *Eco*RI, *Xba*I, *Hho*I, *Hae*III) and fluorescently label the resulting fragments. The total length of the HICF map was estimated 303 Mb (104.5% of the peach genome), which was anchored 45 Mb (15.5%) to the eight linkage groups of the *Prunus* reference map. Peach BACs and physical maps were used, for example, to identify microsynteny regions between plant species like *Prunus*, *Arabidopsis*, *Populus* and *Medicago* with its relevance in the study of evolutionary relationships and candidate gene search (Jung *et al.*, 2006; Jung *et al.*, 2009).

3.4 Whole genome sequencing

A complete genome sequence is a physical map at its maximum resolution (O'Rourke, 2014). Peach whole genome sequencing initiative, coordinated by US and Italian Institutions, was undertaken by the International Peach Genetics Society (IPGI), including the United States (Clemson University and NC State University), Italy (the Italian Consortium funded by the DRUPOMICS project), Spain (IRTA), France (INRA) and Chile (Universidad Andres Bello). Whole genome sequencing, splicing, assembly and gene function annotation was conducted using Sanger whole-genome shotgun sequencing methods in the peach doubled haploid (and therefore homozygous) cultivar 'Lovell' (Toyama, 1974; Germana, 2006).

The first version 'Peach v1.0' was released under Fort Lauderdale Agreement on 1 April 2010 and the results were published in 2013. For high quality chromosome assembly, the scaffolds were processed for chromosomal assignment, anchoring, and orienting aligned with *Prunus* TxE reference map to create pseudomolecules covering each chromosome. The final whole-genome sequence (WGS) assembly contains a total size of 224.6 Mb assigned in 202 scaffolds covering an 8.47× genome depth (Verde *et al.*, 2013). Few years later, peach genome was improved to 'Peach v2.0.a1' with 8 pseudomolecules numbered to the eight peach linkage groups, with enhanced quality of chromosome-scale assembly and annotation of the repeated and gene sequences (Verde *et al.*, 2017). This new version currently consists of a total of 227.4 Mb arranged in 191 scaffolds as well as some repetitive sequences in unmapped scaffolds. Paired-end sequencing WGS Illumina reads were assembled, producing a contig L50 of 255.4 kb (214.2 kb in 'Peach v1.0') and a contig N50 of 250 (294 in 'Peach v1.0'). Altogether, 26,873 protein-coding genes and 28,689 protein-coding transcripts were predicted in the Peach v2.1 annotation (27,852 genes in 'Peach v1.0').

3.5 Use of Peach genetic variability in gene mapping (GWAS)

Genetic variability is the bases for plant breeding. Peach genetic variability is low in terms of allelic diversity (A) and observed heterozygosity (H_o). Aranzana (Aranzana *et al.*, 2003)reported values of an average of number of 7.3 alleles and 0.35 H_o per SSR marker in a collection of 212 peach accessions (commercial and landraces) from US and Europe. Similar values were obtained in the analysis of other peach collections, including mainly landraces (Bouhadida *et al.*, 2010). Such variability values increased when including to the analysis peaches from Asia (mainly China) (A= 12.25 and Ho =0.47) (Li *et al.*, 2013), however when considering both populations (Oriental and Occidental) separately variability values numbers kept low in both of them, indicating that local adaptation and breeding processes have reduced variability in both collections of germplasm. It is notorious that allele frequencies were fixed differentially in each population, indicating that both germplasms are complementary in terms of variability.

In fact, the analysis of the allelic frequencies in germplasm has allowed the identification of the accessions in subpopulations into subpopulations. This information has permitted to infer the extent of linkage disequilibrium (LD) in peach (i.e. to determine at which distance two loci are not inherited independently) and to perform genome-wide association studies (GWAS). GWAS are an alternative to linkage maps. These analysis aims to identify polymorphisms linked to phenotypic measures using populations of unrelated genotypes instead of progenies. This strategy takes advantage of natural genetic variability and of recombination events occurred during generations. GWAS has been broadly used in cultivated species, successfully identifying genomic regions for agronomic traits. In the case of peach, GWAS was used as prove of concept to map major genes already mapped in peach; close to 3,500 peach cultivars from different origins were genotyped with the 9K peach SNP array (Verde *et al.*, 2012) and evaluated for fruit acidity, fruit shape, fruit flesh color, fruit pubescence (peach/nectarine types), fruit texture, flower type (showy/non-showy) and leave gland shape (Micheletti *et al.*, 2015). Similar strategy GWAS was employed with a larger set of SNPs (4,063,377) derived from genome sequencing 129 peach accessions (Cao *et al.*, 2016). The map regions obtained for the 12 agronomic traits where consistent with the ones obtained with linkage mapping, allowing postulating some candidate genes. However, GWAS accuracy depends on the LD extension, which has been described to span about 0.8 to 1.4 Mb depending on the population (Micheletti *et al.*, 2015), limiting the accuracy of this analysis for the inference of candidate genes.

3.6 Somatic variability

Genetic variability is the bases of breeding. When we talk about genetic variability in plants we usually refer to that occurring in germinal cell lines, and, therefore sexually transmitted to the offspring. However, mutations may occur at meristems and therefore such variability will be extended and accumulated into the plant new branches, leaves or flowers. In the case of clonally propagated plants, mutations accumulated in new branches may be maintained and multiplied by grafting, perpetuating the newly generated variability. Consequently, somatic variability is relevant role in species vegetatively propagated. In some cases, novel genetic variability produces novel phenotypes which are called "sports". Numerous examples of sports have been reported in perennial crops like grape, citrus, peach, pear, apple and blackberry, among other (Dermen, 1948; Foster & Aranzana, 2018). Early research by Shamel and Pomeroy listed 1664 bud sports in different fruit tree species, highlighting that at that time 32% of the plant patents issued by the U.S. Patent Office were plants originating as somatic bud mutations. In peach, somatic variability has been broadly exploited. W.R. Okie reports in his 'Hadbook of Peach and Nectarine Varieties' (Okie, 1998) more than 170 cultivars derived from mutations. A careful review of the available information on the origin and pedigree of such cultivars suggests that some varieties are

more prone to generate sports. For example, varieties like 'Redhaven', 'Springcrest' or 'Elberta', among others, have generated more than 40 mutants maintained with more or less commercial interest (Okie, 1998). Recently mutants of the flat peach variety 'UFO4' have been reported to show changes in shape (López Girona, 2014), skin trichrome or flesh color; Figure I.5). Additional proves of the relevance and impact of somatic mutations are the recent publications studying sports and somatic variability (Chatelet *et al.*, 2007; Carbonell-Bejerano *et al.*, 2017), developing bioinformatics tools (Marroni, F. *et al.*, 2017) or reviewing the phenomena (Foster & Aranzana, 2018).
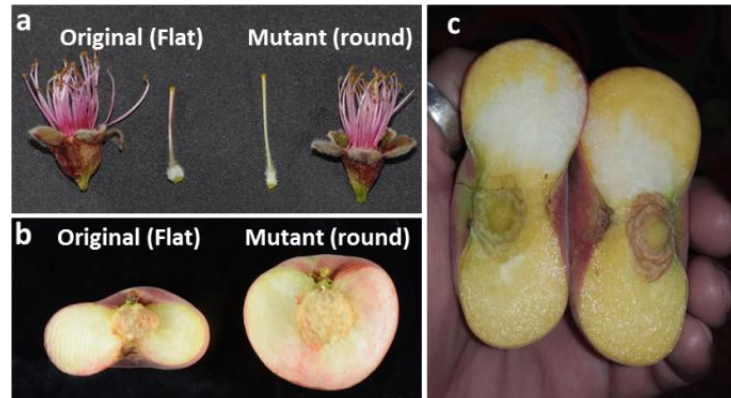
Figure I.5 Somatic variants of 'UFO4' (flat peach with white flesh). Flower and ovary **a**) and fruit **b**) showing flat (original) and round (mutant) shape. **c**) chimerism in flesh color (from white (original) to yellow (mutant)).

Meristems use to consist on several layers of dividing cells and each layer come into different tissues. In peach meristems, the layer I (LI) produces the epidermis (leaf epidermis and fruit skin), the LII layer comes into the outer cortex and part of the vascular cylinder (leaf mesophyll, fruit mesocarp and germ cells in the embryo), and the inner cortex, vascular cylinder result from the LIII (vascular tissues and endocarp of fruit pit) (Dermen, 1953). Usually mutation in meristems affect only one of the layers, therefore the tissues of the new organ are chimeric (i.e. contain original and mutated DNA) resulting, in some cases, in phenotypic mosaicism (Figure I.5 c). Somatic mutations result from changes in DNA sequence or from epigenetic variations. DNA changes are due to errors at replication, recombination and DNA repair during mitosis, which use to produce point mutations, gene deletion or duplication and genome translocations (Foster & Aranzana, 2018). Transposable elements (TE) have shown to play a major role in these events.

It has been extensively described that Transposable elements (TE), especially retrotransposons, are involved in large portion of somatic variation in plants (Lijavetzky *et al.*, 2006; Fernandez *et al.*, 2010; Carrier, G. *et al.*, 2012; Otto *et al.*, 2013). Retrotransposons replicate via reverse transcription of messenger RNAs and integrate into other chromosome locations producing genome rearrangements, disruption of coding sequences or changes in the expression and function of nearby genes. In addition

to retrotransposons, long terminal repeat (LTR) transposons are also involved somatic variability as they can excise due to errors during recombination resulting in genomic loss or chromosome rearrangements. Besides, sports can also be triggered by DNA mutations due to errors during replication, recombination or DNA repair at mitosis. Eukaryotic organisms have developed efficient DNA damage repair mechanisms (DDR) by implicating specialized proteins and regulatory pathways (Manova & Gruszka, 2015), however some of these pathways introduce some DNA changes. For example, non-homologous end joining (NHEJ) and homologous recombination (HR) pathways are involved in the repair of DNA double-strand breaks (DSBs). NHEJ mechanism is likely to occur more frequently than HR (Boyko *et al.*, 2006); the broken extremes ligate without the need of a homologous sequence although may result in small insertions or deletions of DNA at the repaired point. On the contrary, HR pathway needs a template to repair the break as for example the sister chromatid if it occurs in S or G2 phases of mitosis or less frequently the homologous chromosome. While recombination with the sister chromatid will not produce a genetic signal, the use of the homologous molecule will end in a loss of the variability enclosed in the affected region of the broken chromosome which will be replaced with that present in the homologous one (Figure 1.6). This will result in a loss of heterozygosity (LOH) (LaFave & Sekelsky, 2009; Pardo *et al.*, 2009).

In plants, genome rearrangements have been mainly reported in grape. Recently, Carbonell-Bejerano *et al.* described a phenomena involving multiple breaks and rearrangements occurred in a unique event, what is known as chromothripsis (Carbonell-Bejerano *et al.*, 2017). This phenomenon resulted in multiple genetic changes, including regions with LOH, small polymorphisms and lost regions. Chromosome rearrangements are usually demonstrated through the analysis of the genetic sequences as well as by its visualization at chromosome level. One of the complications for the analysis is the frequent chimeric origin of the mutations, which usually affect one of the meristematic layers and, therefore, is not equally present in all plant tissues. Marroni *et al.* developed a bioinformatic pipeline to detect genome rearrangements through the inspection of regions with reduced heterozygosity (ROH) as a result of rearrangements affecting only LII tissues (Marroni, F. *et al.*, 2017).

In addition, chromosome restructuration can be visualized by fluorescent *in situ* hybridization (FISH) technique. FISH is a cytological technique used to localize or detect, by observation in a microscope, the presence of DNA regions in chromosomes hybridized with highly homologous fluorescent probes. This technique has been successfully used in to identify chromosome aberrations and rearrangement principally for human and biomedical research (Bishop, 2010) but also in plants (Szinay *et al.*, 2010). Here, sample preparation and tissue chimerism challenge the success of the technique. The tissue selected for analysis should contain sufficient number of cells in metaphase division (when chromosomes are unfolded), however depending on the tissues used to

obtain the cells and on the cell layer mutated (LI derived cells are much reduced in leaves than those derived from LII) the number of ideal cells may be much reduced.

In summary, the bioinformatics comparison of the whole genome sequence of the original and sport samples followed by chromosome visualization with FISH will provide a determinant tool to describe chromosome rearrangements. However, due to DNA chimerism both methods need to be well designed.

3.7 Use of molecular markers in breeding programs

Peach has an intense variety renewal, with new varieties released yearly to the market. Plant breeding process includes selection of appropriate seed and pollen parents for breeding, perform directed crosses, seed germination, planting and selection during several generations for desired qualities. The use of molecular markers able to predict at seedling stages some of the traits only observable in adult plants (as for example fruit traits) provide a clear benefit for plant breeding, increasing the efficiency as only those seedlings carrying the desired traits will be maintained.

Several markers for major traits are already available and used in public and commercial breeding programs, increasing in importance over years. Public initiatives from USA and from the European Union have put together efforts to bridge the gap between science and breeders, funding projects for such purpose. Among these projects stand out the USA initiative RosBreed (from 2010 and currently in activity) and the European project Fruitbreedomics (2011-2015). The initial joint effort of both projects allowed the design and implementation of markers for fruit morphology traits (peach/nectarine, yellow/white flesh, flat/round shape) as well as for fruit quality traits (acid/low-acid taste, melting/non-melting texture). Currently RosBreed community offers DNA test for some other traits, including resistance to bacterial spot, peach blush, fruit maturity timing and fruit storability (https://www.rosbreed.org/breeding/dna-tests). Also IRTA has a genetic analysis service to aid breeders and growers to identify and characterize molecularly newly released cultivars as well as for MAS.

Molecular markers are also especially interesting for the introduction of desirable alleles from wild or less improved materials into elite cultivars. New breeding strategies have been recently formulated in this sense. Serra *et al.* proposed the MAI strategy to fast exploration and integration of novel variation from exotic resources in long intergeneration species, with the example of almond (exotic resource) and peach (elite material) as prove of concept (Serra *et al.*, 2016). This three steps strategy consists on the generation of a large population and select with markers those with low number of introgressions, phenotyping of this selected lines for selected traits, infer their inheritance and map segregating QTLs, and self or backcross the selected lines with lower number of introgressed fragments. This allows obtaining new improved lines in a

9 - 10 years of period, which is relatively short considering the long intergeneration time of fruit trees.

4   Gene function studies

4.1   Methods and techniques for studying gene function

According to the Gene Ontology (GO) vocabulary system, gene function study can be classified into three perspectives: molecular function (biochemical activity or capability of a gene product, such as enzyme, ligand, *etc.*), biological process (i.e. to which biological objective the gene contributes such as cell growth, signal transduction, *etc*), cellular localization (subcellular structure where a gene product is active, such as ribosome, Golgi apparatus, *etc.*) (Ashburner *et al.*, 2000).

The investigation of a gene function could be tackled by determining the genetic bases of a given phenotype observed naturally or induced (what is called forward or classic genetics) or, in the opposite direction, by studying the effect of induced changes in a particular gene (known as reverse genetics). In Reverse genetic studies, the possible function of the target genes is predicted by the comparison with other gene member of the same family by using bioinformatics tools and/or gene structure information under *in silico.* Reverse genetics have been enormously boosted by the popularization of next generation sequencing techniques, which have contributed to the generation of large amount of sequencing data accompanied by an active research community nourishing databases and developing bioinformatics analysis tools. Such tools (for example Basic Local Alignment Search Tool (blastn and blastx) allow to align newly obtained DNA or amino acid sequences against gene banks to infer a putative gene functions (Johnson *et al.*, 2008). Among sequencing databases, GenBank (Benson *et al.*, 2018) stands out for the large volume of DNA information it contains. Another interesting tool is the online environment PLAZA 4.0, which contains structural and functional gene annotations, homologous gene families, multiple sequence alignments, phylogenetic trees, and co-linear regions within and between species (Van Bel *et al.*, 2018).

To reveal the underlying relationship between the target genes and phenotypic traits, experimental techniques need to be conducted at the cellular level under *in vivo* (or *in vitro*) condition. Therefore, bioinformatics analysis is usually followed by the spatiotemporal *in vivo* gene expression analysis during plant development, for example, what and how mRNA and/or protein expression points and acts in different cell types and developmental phases tested by Northern-Blotting (Western-Blotting for protein analysis), reverse transcription polymorphism chain reaction (RT-PCR), serial analysis of gene expression (SAGE) (Wang *et al.*, 1989; Velculescu *et al.*, 1995). Functional validation methods can be based on the induction of a gain-of-function and on the induction of a

loss-of-function activity. Although both methods imply usually genetic transformation (i.e. obtain transgenic plants), the former involved gene overexpression (include transient expression and stable expression), and gene knock-in, loss-of-function can be achieved by silencing the gene through transformation using RNA-mediated interference (RNAi), endonuclease technology (include zinc-finger nuclease (ZNF), transcription activator-like effector nuclease (TALEN), clustered regularly interspaced short palindromic repeats associated protein (CRISPR/Cas9)), and gene knock-out (Jaenisch & Mintz, 1974; Herskowitz, 1987; Capecchi, 1989; Kim *et al.*, 1996; Werner & Creller, 1997; Boch, 2011; Doudna & Charpentier, 2014). Furthermore, the prevalence of gene cloning, cDNA and DNA microarray, yeast artificial chromosome (YAC), antisense technology, microRNA (miRNA), and gene trapping provides more efficient up-to-date approaches for contributing to the objectives of understanding gene function. The main limitation of these techniques is the difficulties of regenerating transgenic plantlets intrinsic to many species, for example peach.

As an alternative to genetic transformation, plant virus vectors have been exploited as a powerful tool for gene function study through expression of foreign genes into plants (Scholthof *et al.*, 1996; Choi *et al.*, 2001). The usage of this system involves cloning and transferring a foreign gene or a construction into a viral carrier vector. Plants are infected with the recombinant vector by bacterial transformation. Gene expression can result in specific mRNAs degradation of the endogenous plant targeted gen, known virus-induced gene silencing (VIGS) (Burch-Smith *et al.*, 2004), or in the exogenous DNA transcription in the plant nucleus when the transferred gene contains a plant-functional promoter leading to overexpression (Lindbo, 2007). The first attempt of making use of plant virus vector can date back to the R67 plasmid-encoded dihydrofolate reductase (*dhfr*) gene expression based on a double-strand DNA virus *Cauliflower mosaic virus* (CaMV). The final chimeric viral DNA propagated properly in turnip plants by producing a functional enzyme according to *dhfr* ability of regulating methotrexate resistance (Brisson *et al.*, 1984). Nowadays several research works have developed and used successfully different plant virus vectors, such as the potexvirus group members *Potato virus X* (PVX) (Chapman *et al.*, 1992) and *Foxtail mosaic virus* (FoMV) (Mei & Whitham, 2018), the potyvirus *Turnip mosaic virus* (TuMV) (Beauchemin *et al.*, 2005) or *Tobacco rattle Virus* (TRV) (Ratcliff *et al.*, 2008), among other.

4.2 Gene Function and functional validation in peach

Several research works have been addressed to uncover molecular and cytological mechanisms involved in peach fruit development and maturity processes. These studies include gene expression analysis, functional prediction, protein features description, enzymatic analysis, metabolic regulation, cells to cell signal and signal cascade which have contributed to provide an overview of the evolving state of peach genomes as well as of other *Prunus* species.

Most of expression studies reported in peach focus on candidate genes for fruit quality, fruit morphology, stone formation, biotic and abiotic stress resistance, and those involved in biological processes like endodormancy and the requirement of chilling hours. Microarrays (microchips with oligonucleotide probes of unigenes) are used to evaluate the expression levels of a large number of different genes simultaneously. One example is the peach microarray µPEACH1.0 with almost 5,000 probes derived from a set of unigenes expressed at different developmental stages (Trainotti *et al.*, 2006). Hybridization using fruit samples at unripen and ripen stages identified up- and down- regulated genes, most of gene products classified with the genes involved in ethylene biosynthesis and action within the GO annotation for function and biological process, while localized in cell wall and chloroplasts (cellular localization within GO). Microarray analysis identified that transient induction of phenylpropanoid, lignin and flavonoid pathway genes concurred with fruit endocarp lignification and subsequent stone hardening. Among those regulons enriched in cell wall synthesis/modifying genes, 3 were suggested for encoding chloroplast/photosynthesis, 11 for membrane/intracellular transport, 7 for ribosome/ protein synthesis, 9 for proteasome/proteolysis, and 3 for lipid/ fatty acid metabolism (Dardick *et al.*, 2010).

Next generation sequencing techniques have been also successfully applied to the analysis of transcripts. Sequencing the whole RNA (RNA-sequencing) enables to look at expression levels in a more extensive way than microarrays do, providing the whole sequence of the transcripts. In 2013 Wang *et al.* released the transcriptome of peach flowers, leaves and fruit at 2 developmental stages of six peach varieties (Wang *et al.*, 2013). Shortly after, Chen *et al.* used this strategy to analyze differentially expressed genes involved in peach flower variegation. Other differentially expressed genes (DEGs) analysis based on RNA-seq studied expression profiles of genes involved in ribosome, plant-pathogen interaction, flavonoid biosynthesis, and linoleic acid metabolism in peach fruit (Chen *et al.*, 2014). Altogether, 92 DEGs for aroma and fruit softening were clustered in six groups according to the similarity of their expression pattern, of which 78 genes were involved in the pathway of ribosome (ath03010), the most enriched pathway during the fruit ripening process (Li *et al.*, 2015). Other examples of RNA-Seq analysis in peach aim at identifying genes differentially expressed as a response to pathogen infection as *Xanthomonas arboricola* pv *pruni* (Socquet-Juglard *et al.*, 2013) and *Plum pox virus* (Rubio *et al.*, 2015) or to morphophysiological disorders (Bakir *et al.*, 2016), among other. Peach RNA-seq works have been useful to provide information of ESTs and candidate genes, for functional genomics as well as to assist in the peach whole genome annotation.

Despite the large amount of complete information provided by microarrays and RNA-seq methods, differential expression patterns are usually validated by reverse transcription quantitative PCR (RT-qPCR). RT-qPCR enables to determine the amount of

a transcript in a sample through a real time PCR amplification of the DNA derived from the reverse transcription of a particular RNA sample. Therefore, this method is used to analyze the expression of few genes with a known sequence. For example this method has been use to understand the function of NECD genes (9-*cis*-epoxycarotenoid dioxygenase, which encode key enzymes in the abscisic acid biosynthesis) at fruit ripening stages (Thompson *et al.*, 2001; Tan *et al.*, 2003) revealing high expression levels at the beginning of fruit ripening by initiating ABA which preceded ethylene production, which is consequent with the hypothesis that ABA accumulation might account for the regulation of ripeness and senescence (Zhang *et al.*, 2009). Other examples of works conducted to study molecular function of candidate genes involved fruit development and ripening can be found in (Fu & Luan, 1998; Bañuelos *et al.*, 2002; Song *et al.*, 2015) about genes involved in the regulation of K^+ uptake, transport and accumulation during fruit formation and fast growth stages; in (Tonutti *et al.*, 1997; Chang & Stadler, 2001; Alexander & Grierson, 2002) for genes 1-aminocyclopropane-1-carboxylic acid (ACC) synthase (ACS) and ACC oxidase (ACO) gene involved in ethylene biosynthetic pathway.

Bioinformatics analysis of DNA and RNA sequences followed by expression pattern validation through RT-qPCR are first steps to identify candidate genes for a given process. However further *in vivo* validation to confirm the role of such gene(s) is usually required. Although genetic transformation is the most used technique for this purpose, peach is recalcitrant for transformed plantlet regeneration, handicapping the use of this technique. An alternative to genome transformation can be found in the temporary expression of a gene in plant or fruit tissues; this method is known as transient expression analysis. For transient expression, the genes are incorporated into plant cells by agro infiltration or through virus vectors without being incorporated in the plant genome. Examples of transient expression of genes in peach are restricted to those involved in easily observed traits like red coloration caused by MYB10.4 transcription factor (Zhou *et al.*, 2015).

The bibliography reports few VIGs works in peach. One of them, reported by (Li *et al.*, 2017), silenced with this technique the peach homologous to the *Arabidopsis* SEPALLATA gene (*PpSEP1*) by inoculating peach fruit with a construction of *Tobacco rattle virus* (TRV). By inhibiting *PpSEP1* expression, fruit softening of harvested fruit was delayed. A viral vector derived from *Prunus necrotic ringspot virus* (PNRSV) was used to silence the endogenous peach gene *eukaryotic translation initiation factor 4E isoform* (*eIF(iso)4E*), conferring resistance to *Plum pox virus* (PPV) (Cui & Wang, 2016). PPV have been also used as silencing vector in *Nicotiana benthamiana* lines (Vaistij & Jones, 2009); the high susceptibility of some peach lines to PPV suggests that a viral construction with this virus may successfully use to induce the expression or the silence of a gene in peach.

PPV, first described in Bulgaria in 1915 (Atanasoff *et al.*, 1932), belongs to the Potyvirus genus of plant viruses (genus *Potyvirus*; family *Potyviridae*). PPV is a positive-

sense single-stranded RNA virus, with genome size around 10kb (Hollings & Brunt, 1981). Its function is based on several effective helper component proteins (HC), cylindrical inclusion protein (CI), two nuclear inclusion proteinases (NIa and NIb) and the coat proteinase (CP), which is the most variable region among potyvirus polyproteins allowing the discrimination of PPV isolates (Shukla & Ward, 1988; Bousalem *et al.*, 1994). A sixth peptide, the genome-linked proteinase (VPg) is also encoded by the viral RNA (Shahabuddin *et al.*, 1988). By sequencing overlapping cDNA clones of an aphid non-transmissible isolate of PPV (PPV-NAT), Maiss *et al.* were the first releasing the whole nucleotide sequence of its RNA (Maiss *et al.*, 1989; García *et al.*, 2013)**.** Seven PPV strains have been found so far: PPV-C (Cherry), PPV-D (Dideron), PPV-EA (El Amar), PPV-M (Marcus), PPV-Rec (recombinant), PPV-T (Turkey) and PPV-W (Cambra *et al.*, 1994; López-Moya *et al.*, 2000; Glasa *et al.*, 2004). PPV-M is the one infecting peach.

López-Moya *et al.* (2000) successfully developed an infectious construction of a PPV cDNA clone by including its full-length sequences between a cauliflower mosaic virus 35S promoter and a nopaline synthase terminal signal (López-Moya, 2000). This type of constructions , in combination with the green fluorescent protein (GFP) gene (which exhibits a bright green bioluminescence when exposed to light in the blue to ultraviolet range) have been applied to the study virus infectivity, its long-distance movement and the its mechanisms to overcome resistances (Riechmann *et al.*, 1990; Guo *et al.*, 1998; Decroocq *et al.*, 2009). This strategy has also been used in peach. In order to study the interaction between the virus and *Prunus* at the cellular level, Lansac *et al.* developed a GFP-tagged PPV construct which was used to infect the peach rootstock 'GF305' (known to be susceptible to PPC) and the apricot cultivar 'Screara'. GFP fluorescent signal was detected in the medulla and the epidermal cells in 'GF305' (Lansac *et al.*, 2005). Those positive results from *in vitro* grafting inoculation with GFP-tagged PPV provide a useful tool to study virus movement in *Prunus* species.

5   Fruit shape in peach

(Part of this section has been adapted from (López-Girona *et al.*, 2017), which is the first chapter of this thesis.)

Fruits are the edible part of many cultivated species and their study is one of the major topics in plant research. Peach (*Prunus persica* (L.) Batsch) is one of the most economically important fruit species in temperate regions. The fruits are drupes which develop from a single carpel. The calyx and the stamen of the flowers fuse into the hypanthium tissue forming a cuplike structure around the ovary. All peach tissues come from the ovary; the outer skin is the exocarp, the mesocarp the edible flesh and the pit the endocarp. Most peach cultivars are round or oval shaped, although commercial interest in flat shape fruits is increasing fast.

While little is known about the genetic mechanisms regulating fruit morphogenesis in fruit trees, many genetic studies have aimed to unravel such process in the model species *Arabidopsis*. In this species leucine-rich receptor like kinases like *ERECTA* and *CLAVATA-1*, show functional implications in the maintenance, size and shape meristem (Torii *et al.*, 1996; Mandel *et al.*, 2014). In particular, *ERECTA* regulates organ shape and flower architecture, showing the loss-of-function *ERECTA* mutants compact inflorescences, short pedicels and round flowers (Torii *et al.*, 1996; Shpak, 2003).

Among cultivated species, fruit shape has been most studied in tomato. The fruits are berries which develop from the ovary after fertilization of the ovules. The wall of the ovary develops into the pericarp and encloses the placenta and seeds. Four genes controlling tomato fruit shape have been cloned: *SUN* (Xiao *et al.*, 2008), *OVATE* (Liu *et al.*, 2002), *LOCULE NUMBER* and *FASCIATED* (Muños *et al.*, 2011; Rodriguez *et al.*, 2011). In addition, several loci which regulate fruit shape have been identified including two suppressor elements of the ovate mutation (*Sov1* and *Sov2*) (Rodríguez *et al.*, 2013). One is the mutant *Self1*, producing fruit elongation by increasing cell layers in the ovary (Chusreeaeom *et al.*, 2014), and the other is QTL *fs8.1* which also controls fruit elongation (Paran & van der Knaap, 2007). *SUN*, *OVATE*, and *fs8.1* act together in additive manner to control fruit shape producing longer fruits. In cucumber, a homolog of the tomato *SUN* gene (*CsSUM*) is a candidate for round fruit shape (Pan *et al.*, 2017).

Flat peaches originated in South China, where they are known as "pentao" from the original Chinese "Pan Tao". In the mid-1800s several Chinese flat varieties were introduced into USA breeding programs as carriers of characters such as low chilling [13], but they were popular for a brief period of time. It is believed that the first flat peach variety, bred by Starks Nursery in 1985, was 'Saturn' and later, in the 1990s, its cultivation became more widespread (Layne & Bassi, 2008). The flat shape of the peach fruit is determined in the early stages of flower development (Figure I.6) by a single gene *S/s* (for saucer-shaped) mapped in the distal part of chromosome 6 (Dirlewanger *et al.*, 1998). Fruit from individuals with the *ss* genotype is round, those heterozygous for the flat allele (*Ss*) are flat, and fruit from homozygous *SS* plants abort several weeks after anthesis. Although the hypothesis of a single gene explains the phenotypes observed, abortion of homozygous *SS* plants also suggests two dominant closely linked genes (*S/s* and *Af/af*) in repulsion (Dirlewanger *et al.*, 2006). Up to now, several markers have been identified around the *S* locus, by analyzing both mapping progenies and germplasm (Picañol *et al.*, 2013; Cao *et al.*, 2016; Lambert *et al.*, 2016). One of the markers, the SSR UDP98-412 has been reported to be tightly linked to the *S* locus and works efficiently in marker assisted selection (MAS) (Picañol *et al.*, 2013).
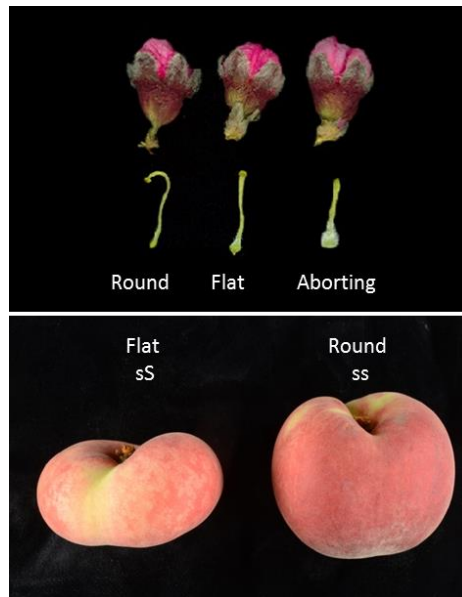
Figure I.6 Flowers and ovaries of round, flat and aborting types, and flat and round fruit with the corresponding genotypes.

Horn *et al.* mapped ESTs of 3,842 candidate genes for fruit quality in the *Prunus* reference map, but no candidate genes were identified for fruit shape (Horn *et al.*, 2005). Recently, a *PpCAD1* gene (*Prupe.6G292200*, alias *ppa003772m* peach genome v.1) has been reported as candidate for the trait based on a GWAS analysis (Cao *et al.*, 2016). However, the expression pattern of this gene does not agree with the trait determination during fruit development and therefore more studies are required in the identification of a candidate gene for flat shape in peach.

The *S* locus determined by linkage mapping (approximately Pp06:26.217.000.. 26.617.000) contains 115 genes. The most represented gene family is that of genes encoding receptor-like protein kinases (RLKs), with 5 genes (*Prupe.6G281100*; *Prupe.6G281200*; *Prupe.6G281300*; *Prupe.6G281400*; *Prupe.6G281500*) clustered in a 23 Kb region. RLKs constitute a large superfamily of proteins with similar structure which are involved in diverse plant processes like development, growth and response to pathogens. They consist of an extracellular region, a single-pass transmembrane spanning domain, and an intracellular kinase domain. When extracellular small signaling molecules (ligands) physically bind with the extracellular region, two RLK molecules join to generate homo or hetero-dimers prompting the phosphorylation of both RLK molecules and therefore transducing the signal to downstream networks altering gene transcription patterns (Li & Tax, 2013).

Plant RLKs are serine/threonine protein kinases, which are classified into subgroups based on amino acid sequence motifs upon the extracellular domains: S-domain class (S-RLKs), leucine-rich repeats class (LRR-RLKs), tumor-necrosis factor receptor-like repeats class (TNFR), the epidermal growth factor-like repeats (EGF),

pathogenesis related protein 5-like receptor kinase class (PR5-RLKs), lectin class (lectin-RK), and other RLKs sharing no homology to known motifs (Walker, 1994; Braun & Walker, 1996; Hervé *et al.*, 1996; Wang *et al.*, 1996; McCarty & Chory, 2000). Among these groups, LRR-RLK, containing a tandemly repeated motif with conserved leucine, is the largest group of plant RLKs family with more than 200 members (Shiu & Bleecker, 2001).

Currently, the LRR-RLKs are commonly categorized into 13 subfamilies (LRR I to LRR XIII) based on the amino acid relationships between their kinase domains (Kobe & Deisenhofer, 1994; Shiu & Bleecker, 2001). A typical LRR-RLK motif is composed of 23 – 25 amino acids as well as several β-α structural LRR units with a consensus sequence pattern 'Leu-x-x-Leu-x-Leu-x-x-Asn-x-Leu' in the extracellular domain for ligand binding (Kobe & Deisenhofer, 1995). LRR-RLKs, thought to be involved in protein-protein interaction, have been found to function in a variety of plant processes where some LRR-RLKs are involved in regulating the tissues growth and development. In *Arabidopsis*, *SERK1/2* (somatic embryogenesis receptor-like kinase), *RPK2* (receptor-like protein kinase), and *EMS1* (express microsporocytes) were identified to regulate microsporogenesis and embryogenesis development (Albrecht *et al.*, 2005; Zhou *et al.*, 2016). The *CLV* (*CLAVATA1/2/3*) family linked with cell proliferation and differentiation in the shoot apical meristem, while the homeobox transcription factor *WUS* (*WUSCHEL*) which was a downstream target of *CLV1* signal formed a negative feedback loop with *CLV3*, failing to maintain organogenesis (Kayes & Clark, 1998; Sablowski, 2007). Besides, *ERECTA* regulates organ shape and inflorescence architecture, *HAESA* controls floral abscission process, *RUL1* (*REDUCED IN LATERAL GROWTH1*) and *MOL1* (*MORE LATERAL GROWTH1*) are involved in secondary growth (Torii *et al.*, 1996; Jinn *et al.*, 2000; Shpak, 2003; Agusti *et al.*, 2011). Some LRR-RLKs play role in abiotic and biotic stresses, such as *FLS2* (*FLAGELLIN SENSITIVE 2*), found to be in bacterial flagellin recognition and resistance; the rice gene *Xa21* was identified to confer specific resistance to *Xanthomonas oryzae* pv. *oryzae* race 6; tomato *NIKs* (NSP-interacting kinases) mediates antiviral responses (Song *et al.*, 1995; Gómez-Gómez & Boller, 2000; Fontes *et al.*, 2004). Moreover, many LRR-RLKs show a perception ability and transcriptional response to the plant hormones. *RPK1* (receptor-like kinase 1) functions in ABA (abscisic acid) signal transduction pathway and accounts for early ABA perception in *Arabidopsis*; *PSKRs* (phytosulfokine receptor kinases) act as a ligand-receptor pair involved in perception of the peptide hormone phytosulfokine; *BRI1* (brassinosteroid-insensitive 1) and *BAK1* (*BRI1*-associated receptor kinase 1) trigger brassinosteroid (BR) perception to allow downstream signaling of *BES1* (*BRI1 EMS* suppressor 1) and *BZR1* (brassinazole resistant 1) transcription factors, positively regulating a BR-dependent plant growth pathway, and negatively regulating a BR-independent cell-death pathway (Matsubayashi *et al.*, 2002; Osakabe *et al.*, 2005; He *et al.*, 2007; Aker & de Vries, 2008).

Non-functional alleles of LRR-RLKs may show a range of severity effect (Clark *et al.*, 1993), probably due to truncated receptor, with may produce a dominant-negative effect (Shpak, 2003). In the other hand functionally redundant RLK signaling pathways are required to fine-tune the proliferation and growth of cells in the same tissue type during organ development.

Objectives

This thesis aims at the identification and characterization of candidate gene(s) for fruit shape trait in peach. To achieve this goal, we focused on the following specific objectives:

1. Explore genetic diversity in the *S* locus and its linkage with peach fruit shape trait to:

    1.1 Identify genetic variants linked to the fruit flat shape variability.
    1.2 Identify candidate(s) gene(s) involved in fruit shape determination.
    1.3 Understand genetic determination of fruit flat shape in peach.

2. Study genetic variability between a flat variety and its sport mutant which exhibits different fruit shape.

3. Initiate the development of protocols in peach to:

    3.1 Visualize chromosome rearrangements in somatic mutant cells in FISH technology.
    3.2 Validate gene function by inducing its expression in plants through a virus vector.

This thesis is structured in four chapters. Chapter 1 (A deletion affecting an LRR-RLK gene co-segregates with the fruit flat shape in peach) and Chapter 2 (characterization of genetic variability in LRR-RLKs clustering in the *S* locus) approaches objective 1. Chapter 3 (Genetic characterization of a flat variety and its sport mutant which exhibits different fruit shape) tackles objectives 2 and objectives 3.1, while Chapter 4 (PPV virus-based construction for gene function validation) addresses the objective 3.2.

Most of Chapter 1 work has been already published in (López-Girona *et al.*, 2017), therefore part of the published text has been included in this manuscript.

Chapter 1: A deletion affecting an LRR RLK gene co-segregates with the fruit flat shape trait in peach

Elena López Girona[1], Yu Zhang[2], Iban Eduardo[1], José Ramón Hernández Mora[1], Konstantinos G Alexiou[1], Pere Arús[1], María José Aranzana[1]

[1]IRTA (Institut de Recerca i Tecnologia Agroalimentàries), Barcelona, Spain.
[2]Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Barcelona, Spain.
Correspondence and requests for materials should be addressed to M.J.A. (email: mariajose.aranzana@irta.cat)

1.1 Materials and methods

1.1.1 Plant material

In total we studied 249 peach individuals, classified as round, flat or aborting (in those cases where either all or most of fruit set stopped within a few weeks after pollination). Of these, 177 corresponded to peach cultivars (110 round and 67 flat; see Supplementary Table 1.S2). Seventy-one were F1 seedlings from the cross between the two flat peaches 'UFO-3' × 'Sweet cap' (with round (14), flat (40) and aborting (17) phenotypes) and three were round, flat and aborting seedlings (P07F202A065, P07F202A071 and P07F202A056, respectively) from 'ASF08.81' open pollination. In addition, we included a flat peach variety ('UFO-4MUT) and its round somatic mutant ('UFO-4Mut'). Buds of the branch containing the mutation were grafted and maintained at the greenhouse facilities of IRTA at Torre Marimon (Barcelona).

1.1.2 DNA and RNA extraction

DNA from all materials used was extracted from young leaves using the Doyle and Doyle method (Doyle, 1990). For 'UFO-4MUT and 'UFO-4Mut' DNA was extracted from leaves, flesh fruit, skin and stone using the DNAsy Qiagen kit (Qiagen, Hilden, Germany). Branches with flowers in Baggiolini stage E (not expanded petals) from round (P07F202A065), flat (P07F202A071) and aborting (P07F202A056) peaches (all progenies from 'ASF08.81' open pollination) were cut in the field, and pistils collected and frozen in liquid nitrogen and conserved at -80 °C prior to total RNA extraction using the RNeasy® Plant Mini Kit (Qiagene) following the manufacturer's protocol. RNA integrity was confirmed by 1% agarose gel electrophoresis.

1.1.3 DNA genotyping to identify new polymorphisms associated with the trait

All samples were genotyped with the SSR marker UDP98-412 (Pp06: 26,617,638..26,618,013) using the PCR and electrophoresis conditions described in (Picañol *et al.*, 2013).

Using the peach genome sequence v.2 (Verde *et al.*, 2017) we designed 23 primer pairs to amplify fragments of 200-700 bp in a 388.6 kb region (Pp06: 26,254,140..26,642,759) (Supplementary Table 1.S1). Fourteen of them (primers UDP98-412(−17 K) to UDP98-412(+25 K)) were designed covering a 42.8 Kb region (Pp06: 26,599,970..26,642,759) flanking UDP98-412 and the nine remaining (primers Amplicon 1 to Amplicon 9) in a region spanning 26.6 Kb (Pp06: 26,254,140..26,280,026) 363.5 Kb upstream. This region was the closest to UDP98-412, where SNPs of the 9 K peach chip

(Verde *et al.*, 2012) were identified. Primers were designed using Primer3 software (Rozen & Skaletsky, 1999) avoiding amplification of SSR regions.

Primers were first tested in six varieties, three with flat ('Mesembrine', 'Paraguayo delfín' and 'Subirana') and three with round fruit ('Garcica', 'HoneyGlo' and 'Luciana'). PCR products with a single band were purified with Exosap-it (GE HealthcareLife Science) in a single pipetting step and used as template for sequencing using the BigDye^TM Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and forward primers. The sequencing reaction profile included an initial denaturation at 96 °C for 1 m, followed by 25 cycles of 96 °C for 10 s, 50 °C for 6 s, and 60 °C for 4 min; the sequences obtained with an ABI Prism 3130 × l DNA Analyzer (Applied Biosystems, Foster City, California, CA, USA) were visualized and manually edited with Sequencher 5.0 software (Gene Codes Corporation; Ann Arbor, MI, USA). Fragment ends were trimmed to remove low-quality sequence. Haplotypes were graphically represented with Flapjack software (Milne *et al.*, 2010).

1.1.4 Cloning of PCR fragments

For one flat variety ('UFO-8') PCR products were cloned into the pGEM T-easy vector (Promega) following the manufacture instructions. *Escherichia coli DH5α* electro competent cells (Invitrogen) were transformed with the ligated plasmid by electroporation in the Gene PulserXcel electroporation system (BIORAD), with a capacitance 25 μF, resistance of 200 Ω and a voltage of 1,8 Kv. Transformed cells were shaken horizontally at 250 rpm and 37 °C for 1.5 h in 1 ml liquid Luria-Bertani (LB) medium. Fifty microliters of transformed cell solution were then pipetted onto 10 cm LB agar plates containing 50 μg/mL ampicillin, 80 μg/mL X-gal and 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Positive colonies were picked from the LB plates as template DNA for colony PCR. Colonies were screened by PCR following the conditions described above. Those carrying the desirable allele were grown in 5 mL of LB liquid broth containing 50 μg/mL of carbenicillin with overnight incubation at 37 °C in a shaking oven at 250 rpm. Bacterial culture pellets were obtained by centrifugation at 3000 rpm for 10 min. Plasmids were extracted from bacterial cells using a QIAprep miniprep spin-kit (Qiagen) according to the manufacturer's protocol, then resuspended in 50 μL of sterile water and 4 μL of each extract were sequenced with the vector specific primers, either T7 or SPS6, following the sequencing protocol previously described

1.1.5 Sequencing *Prupe.6G281100*

For the round shape-associated allele, using the peach genome sequence as reference, we designed seven primer pairs (Supplementary Table 1.S4) flanking and within *Prupe.6G281100* (Pp06:26,269,777..26,272,029) to obtain the full sequence of the

gene. Primers were designed to amplify single fragments, avoiding amplification of duplicated regions. Amplification and sequencing reactions were as described above.

For the flat shape-associated allele, the forward primer *Prupe.6G281100* (−10K)_F was designed 10,072 bp upstream *Prupe.6G281100* and was combined with the reverse primer *Prupe.6G281100*_3PrimF (primer combination 2 (PC2) in Supplementary Table S4), 558 bp downstream of the gene, to amplify fragments with an expected size of 12.9 Kb. For long-range PCR, LongAmp® Taq Polymerase (New England BioLabs ®INC) was used. Each reaction contained 1x LongAmp reaction buffer, 0.3 mM dNTP mix, 0.8 μM each primer, 5% DMSO, 5 units of polymerase, 40 ng of template DNA, and sterile Milli-Q water to a final volume of 25 μl. The following PCR protocol was performed on a S-1000TM Thermal Cycler (Bio-Rad Laboratories, Inc.; Hercules, California, USA): 95 °C for 5 min; 35 cycles of 95 °C (30 sec), 60 °C (30 sec) and 65 °C (17 min); followed by a final step at 65 °C for 10 min. All PCR amplicons were checked on 1% agarose gel in TAE buffer. Ethidium bromide staining was used for band visualization.

The PCR bands were purified with the High Pure PCR product purification kit (Roche Diagnostic, Basel, Switzerland). Thirty Nano grams of purified product were used as template to obtain the whole sequence of the amplicons in four sequencing reactions using the primers *Prupe.6G281100*(−10K)_F, *Prupe.6G281100*_4 R, *Prupe.6G281100*_5 R and *Prupe.6G281100*_3PrimF (Supplementary Table S4).

1.1.6 Variant validation with NGS

To validate the large variant alignment, we re-sequenced, with Illumina technology (27×), five flat ('Flatmoon', 'Cakereine', Blanvio-10', 'Subirana', 'UFO-4MUT) and five round ('Nectalady', 'Armking', 'Belbinette', 'Nectaross', 'Tifany') varieties. High quality DNA of each sample was delivered to the CNAG (Centre Nacional d'Anàlisi Genòmica, Barcelona) for library preparation and 2 × 100 bp paired-end sequencing using illumina HiSeq. 2000 sequencer. Adapter removal and quality-based trimming of the raw resequencing data was with Trimmomatic version 0.36 (Bolger *et al.*, 2014). FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) was used for read quality control before and after trimming. High quality reads were mapped to the peach genome version 2.0 using BWA (Li & Durbin, 2009) and the resulting alignment files were sorted and filtered by dis- carding multi-mapped reads and annotating PCR duplicates. Reads mapping to the Pp06:26,257,000..26,273,000 region were extracted from the alignment files and bulk aligned against the reference peach genome v2 using CLC-genomics workbench 8.5.1 (https://www.qiagenbioinformatics.com/). CLC-InDels and Structural Variants analysis tools were run separately in the flat and round peach alignments. Genes annotated in v.2 were downloaded from the GDR database (Jung *et al.*, 2014) and included in the alignment track for visualization. Reads are available at the European Nucleotide Archive under the accession number ENA: PRJEB21538.

1.1.7 Design of markers for genotyping

To validate the polymorphisms in germplasm and progenies, we designed primers to amplify a small INDEL within the candidate gene as well as the large deletion upstream from the candidate gene. Thus, primer pairs FlatIn_F and IndelS_R (both inside the gene; PC4 in Supplementary Table S4) yielded product sizes of 464 bp and 469 bp for the flat and round alleles, respectively. PCR conditions, fragment separation and analysis in the ABI Prism 3130 × l DNA Analyzer were as previously described for the SSR marker.

A three-primer combination (PC3 in Supplementary Table S4), consisting of two primers flanking the deletion (one forward and one reverse) and an inner reverse primer (IndelS_F + IndelS_2 R + IndelS_R), was designed to genotype the large deletion identified in this region; IndelS_F and IndelS_R (flanking the deletion) amplified a 1,620 bp fragment associated to the flat phenotype, while IndelS_2 R (within the deletion) in combination with IndelS_F produced a 941 bp band associated to the round phenotype. PCR was carried out in a 10 µl reaction containing 20 ng of DNA, 1x PCR buffer, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.2 µM of each primer and 1 U of BIOTAQ (Biolab). The following PCR protocol was used in a S-1000TM Thermal Cycler (Bio-Rad Laboratories, Inc.; Hercules, California, USA): 95 °C for 2 min; 35 cycles of 94 °C (30 sec), 55 °C (30 sec), 72 °C (60 sec); followed by a final step at 65 °C for 10 min. All PCR amplicons were checked on 1% agarose gel in TAE buffer. Ethidium bromide staining was used for band visualization.

1.1.8 RT-PCR analysis

RNA was reverse-transcribed to cDNA using the reverse primer IndelS_R (see Supplementary Table S4). For this, 1 µL of RNA was hybridized with 2 µL of primer in a total volume of 13 µL. After 10 min incubation at 70 °C and 5 min cooling on ice, cDNA was obtained using PrimeScript RT-PCR kit (Takara). PCR was conducted with PC4 following the protocol described above. The forward primer was florescent labeled to check the size of the fragment in the ABI Prism 3130 × l DNA Analyzer.

1.1.9 Gene homology and functional prediction

Functional annotation and orthologues for the PRUPE.6G281100 gene were determined using Dicots PLAZA 3.0 (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v3_dicots/) (Proost *et al.*, 2015). Custom DNA BLAST (blastn program) against PLAZA Transcript Sequences database were used for similarity searches, filtering for low complexity and using the BLOSUM62 score matrix.

For the protein sequence of *Prupe.6G281100*, associated with the round allele, the DNA sequence was entered in the Translate tool of the ExPASy Bioinformatics Resource

Portal (http://web.expasy.org/translate/). Similarity searches were performed on the NCBI web page (www.ncbi.nlm.nih.gov) against the nr (non-redundant) collection of sequences in GenBank and the UniProtKB/SwissProt databases, using the blastp and the Position-Specific iterated BLAST algorithm (Altschul *et al.*, 1997). The quality of the pairwise sequence alignment was evaluated in a BLOSUM62 protein substitution matrix allowing a gap existence value of 11 and an extension value of 1.

1.2 Results

Search for polymorphisms associated to the flat shape trait. To find DNA polymorphisms associated with the flat trait in peach, we explored a 30.2 Kb region flanking the SSR UDP98-412, previously reported to be tightly linked to this trait (Picañol *et al.*, 2013). We designed 14 primer pairs to amplify and sequence fragments of 350 - 680 bp in this region, in a small set of three flat and three round peaches. No polymorphisms (SNPs or INDELs) were observed in the DNA fragments amplified by these primers.

The SNPs closest to UDP98-412 annotated in the peach genome database occur 363.5 Kb upstream of this marker, with 20 SNPs in a 26.5 kb region (Pp06: 26,254,140..26,254,809). All these SNPs were located in the coding regions of five annotated transcripts. By sequencing nine amplicons of these transcripts we confirmed these 20, plus 10 additional, SNPs in the same set of six flat and round peaches. Thirteen out of the 30 SNPs were associated with the flat phenotype in the small panel of cultivars. All 13 SNPs occurred in a total of 1,150 bp of two partially overlapping amplicons, nine in Amplicon5 and four in Amplicon6 (Supplementary Table 1.S1). In addition, we detected an insertion/deletion (INDEL) polymorphism in Amplicon5 in heterozygosis in flat varieties. To con- firm the association of the SNPs and the INDEL with the phenotype we sequenced Amplicon5 and Amplicon6 in 112 varieties (65 round, 47 flat) and three aborting phenotypes (Aborting02, Aborting08 and Aborting17) from the 'UFO3' × 'SweetCap' progeny. All round varieties were homozygous for the reference allele in 11 out of the previous 13 SNPs and the flat ones heterozygous, while the aborting seedlings where homozygous for the alternative allele, in agreement with the genetics of the trait (Supplementary Table 1.S2). Alignment of the round and aborting sequences of Amplicon5 against the reference genome gave two INDEL variants in the aborting sequence: an 8 bp insertion and, a few bases downstream, a 13 bp deletion. Forward and reverse sequences of Amplicon 5 in flat varieties revealed that they contained both INDELs in heterozygosis. By cloning and sequencing the PCR product of one flat variety ('UFO-8') we confirmed that each of the two alleles were identical to round and aborting, respectively. The two haplotypes observed for Amplicon5 and Amplicon6, in homozygosis or heterozygosis in the small panel of flat and round varieties, are shown in Figure 1.1.
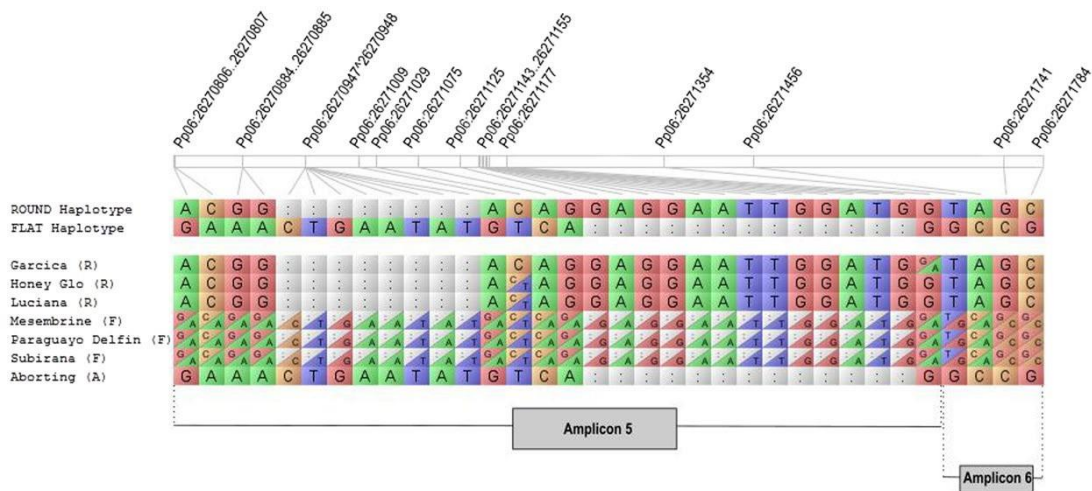
Figure 1.1 Round and flat associated haplotypes in round (R), flat (F) and aborting (A) peaches. Clone represents the deletion of a nucleotide. The haplotypes consist of 13 SNPs and two INDELS, in Amplicon5 and Amplicon6 (Pp06: 26,270,679.26,271,829) (see Supplementary Table 1.S1).

According to the genome annotation, Amplicon5 and Amplicon6 are part of coding regions of a 2,253 bp long transcript (*Prupe.6G281100*). Custom DNA BLAST analysis with PLAZA 3.0 gave significant alignment of *Prupe.6G281100* with seven genes in the peach genome; four (*Prupe.6G281100*, *Prupe.6G281200*, *Prupe.6G281300* and *Prupe.6G281400*) in a region of 36.4 Kb in chromosome 6 containing *Prupe.6G281100*; one in chromosome 7 (*Prupe.7G088700*) and the two remaining in chromosome 8 (*Prupe.8G054400* and *Prupe.8G054300*).

1.2.1 Variant identification and allele cloning

To obtain the whole sequence of the flat and round associated alleles of *Prupe.6G281100* we used primers upstream, downstream and within this gene in two round ('Garcica' and 'Honey Glo'), two flat ('Paraguayo Delfín' and 'Mesembrine') and two aborting individuals. Amplification with the primer pair flanking the gene (PC1) gave a fragment with the expected size (3.3 Kb) in the round and flat DNA, but failed with the aborting ones (Figure 1.2 a). The sequence of the amplified fragment in the flat varieties revealed the presence of the round allele only (lacking the two INDELs). The failed amplification of the flat-associated allele suggested a polymorphism in or near the gene.

To explore this hypothesis, we amplified the samples with forward and reverse primers designed at opposite ends of the gene. Primers 10,072 bp upstream and 558 bp downstream of the gene (PC2) yielded one band of the expected size for the round sample (12,882 kb), and one about 10 Kb shorter in the aborting and in the flat peaches (Figure 1.2 b). The full sequence of the short band revealed a fragment with 2,912 nucleotides long, and consequently 9,970 bp less than that expected from the reference genome. The polymorphisms consisted in the loss of a region from 9,324 bp upstream

of the start codon (Pp06: 26,260,453) to 693 bp downstream of this codon (Pp06: 26,726,336). Despite amplifying the large band in the round samples, where it occurred in homozygosis, we were not able to obtain this band in the heterozygous flat samples, where the short allele appeared to amplify preferentially. As a result, the presence of this fragment in heterozygosis was validated with a three-primer PCR assay (PC3): two (IndelS_F and IndelS_R) flanking the deletion to amplify a 1,620 bp associated to the flat phenotype and one internal to the deletion (IndelS_2 R) to produce a 941 bp band associated to the round phenotype (Figure 1.2 c and d).
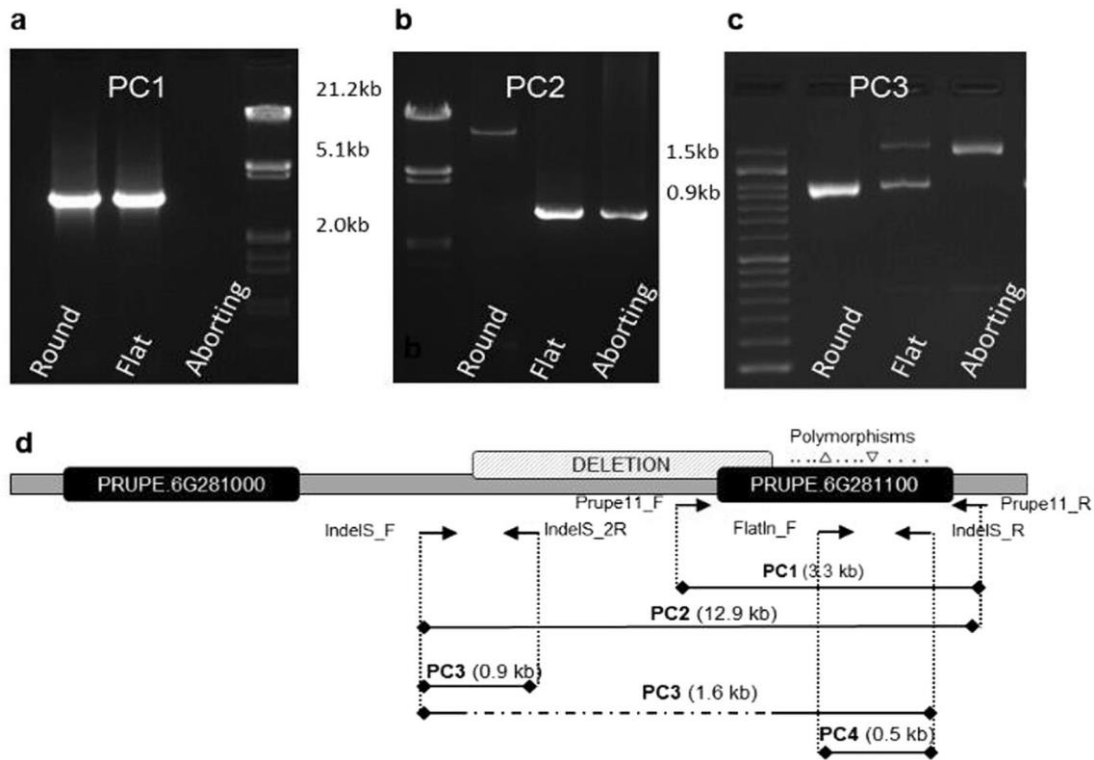
Figure 1.2 PCR bands reveal the deletion affecting the gene Prupe.6G281100 in aborting and flat peaches. **a)** PC1 failed to amplify the flat-associated allele. **b)** Long-range PCR amplification with PC2 produced a fragment about 10 kb shorter in flat and aborting than in round peaches. **c)** PCR-amplification with PC3 identified round (941pb), flat (1620/941 bp) and aborting (1620 bp) genotypes. **d)** Diagrammatic representation of the position of the primers used to identify the polymorphisms associated with flat shape and the polymorphisms (SNPs and small INDELs) in Prupe.6G281100, represented as dots and triangles (respectively).

1.2.2 Variant validation with NGS

The presence of the large deletion in heterozygosis in flat varieties was also validated by resequencing five flat and five round varieties with Illumina NGS technology. The alignments corresponding to Pp06:26,257,000..26,273,000 region (16 kb) of varieties of each fruit type were bulk-analysed. With the CLC-INDELs and Structural Variants tool, the two small INDELs within the gene were identified, but not

the large deletion. However, this deletion was evident in the visual track of the alignment, where only flat peaches had less reads in the deleted region (Figure 1.3).
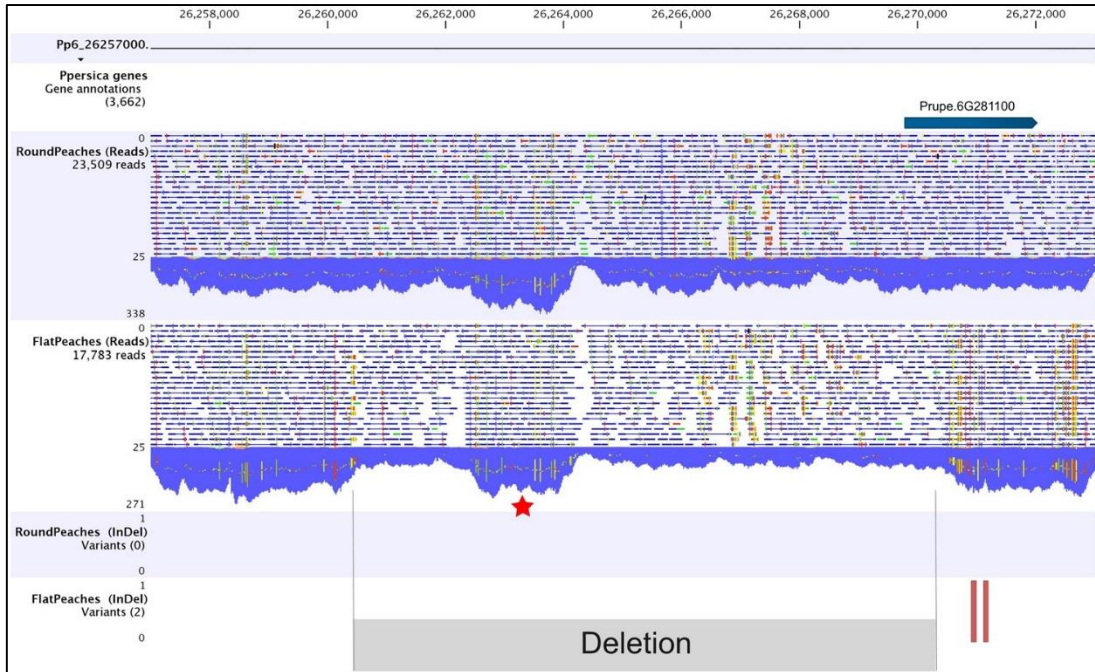
Figure 1.3 Alignment of round and flat peaches reads against Pp06:26,262,400..26,264,250 region. CLC-Workbench track display including **a)** Pp06:26262400..26264250 region Prupe.6G281100, **b)** bulked alignment of Illumina reads from five round and **c)** five flat peaches. Blue areas represent the sequence depth at each position. The reduction in the number of reads in the flat peaches reveals the deletion in heterozygosis. An increase in the number of reads in the region Pp06:26,262,400.26,264,250 (labeled in the figure with **a***) is produced by the spurious alignment (confirmed by Sanger sequencing) of a highly repetitive region. **d)** The CLC- InDels and structural variants tool identified the two indels in the flat varieties only.

1.2.3   Polymorphism validation in peach germplasm and markers for seedling selection

The two small INDELs within the gene and the close to 10 kb deletion were tested with PC4 and PC3, respectively, in a panel of 177 flat, round and aborting samples (Supplementary Table 1.S2). All genotypes matched the observed phenotype. For the two small INDELs, the size of the fragments (469 bp for the round and 464 bp for the flat-associated alleles) confirmed that the two INDEL variants were in heterozygosis in flat varieties and seedlings while the respective alleles were homozygous in round and aborting ones. Similarly, the genotype obtained with the primers flanking and within the 10 kb deletion matched the phenotype (941 bp for the round and 1,620 bp for the flat associated alleles).

1.2.4 Expression analysis

RT-PCR amplification of RNA extracted from pistils of round, flat and aborting peaches produced fragments exclusively from the round and flat samples (Figure 1.4). The size of the bands and their sequence revealed that, in both cases, the fragment amplified corresponded to the round-associated allele (lacking the two small INDELs), indicating the absence of transcription of the flat-associated allele.
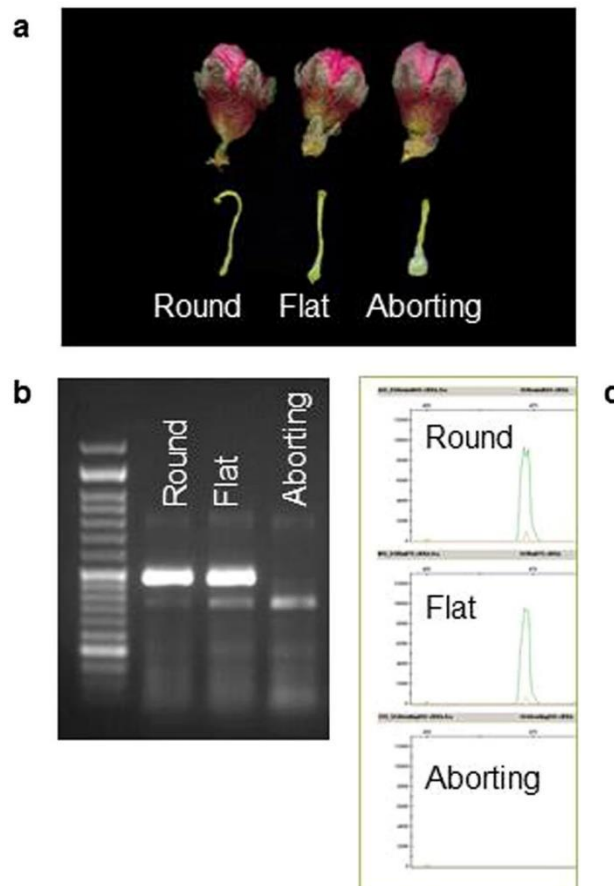
Figure 1.4 RT-PCR of RNA from round, flat and aborting pistils. **a)** Pistil shape observed in flower buds in stage E. On RT-PCR amplification of round, flat and aborting pistils using PC4 no amplification of the flat-associated allele was visible in both **b)** agarose and **c)** capillary electrophoresis.

1.2.5 Homology and functional prediction of the gene

The gene *Prupe.6G281100* (alias *ppa025511m* in the peach genome annotation v.1) is a leucine-rich repeat kinase (PLAZA 3.0 gene family HOM03D000009, orthologue group ORTHO03D000261 described as *BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1, BAK1*). BLAST analysis of the translated protein against the PLAZA protein sequence database gave best alignments with 250 genes from 55 subfamilies of the gene family HOM03D00009 (Supplementary Table 1.S2). Most of the genes (160

genes, 64%) belonged to four subfamilies (ORTHO03D000261, ORTHO03D000539, ORTHO03D001987, ORTHO03D002896). Thirty-eight of them were annotated as *BAK1* (23.8%), 30 (18.8%) as leucine-rich repeat receptor-like kinase protein *FLORAL ORGAN NUMBER1* and 81 (50.6%) as leucine-rich repeat receptor–like kinase protein *THICK TASSEL DWARF1* in 25 species, including peach and other *Rosaceae* members such as *Fragaria vesca* and *Malus* x *domestica*. Within peach, these were *Prupe.6G281000, Prupe.6G281200, Prupe.6G281300, Prupe.6G281400, Prupe.6G281500, Prupe.6G288800, Prupe.7G088700, Prupe.8G054400* and *Prupe.8G054300.*

Protein BLAST pairwise alignment against the SwissProt Arabidopsis database gave significant hits with LRR-RLK, involved in different biological processes. Best hit occurred with the *AtRLP12* gene. This gene may functionally complement *CLAVATA2*, a key regulator that acts at the shoot apical meristem (SAM) of plants, controlling the stem cell population size (Wang *et al.*, 2010).

1.2.6 Analysis of the polymorphism in a round somatic mutant

We analyzed the polymorphisms in a round peach generated from a somatic natural mutant of the flat variety 'UFO-4MUT (Figure 1.5). The analysis of genomic DNA with PC4 (with forward and reverse priers flanking the two small INDELs inside *Prupe.6G281100*) showed a faint amplification of the flat allele in the mutated round cultivar compared to the strong signal observed in the original flat. Amplification of DNA extracted from skin, flesh and stone tissues revealed the absence of the flat associated allele in the flesh mutated DNA while it was present in the skin DNA, indicating that the mutation occurred in the meristematic LII. Faint amplification of the flat allele was observed in the stone DNA of the mutant, which could be due to the invasion of LIII by mutated LII cells.

As for 'UFO-4MUT, amplification of 'UFO-4Mut' flesh DNA with the primers flanking *Prupe.6G281100* (PC1) produced only the round-associated allele, as occurred with those designed to genotype the 10 kb deletion (PC3). These results suggest a mutation affecting the flat associated allele which could have caused the reversion of the phenotype from flat to round.
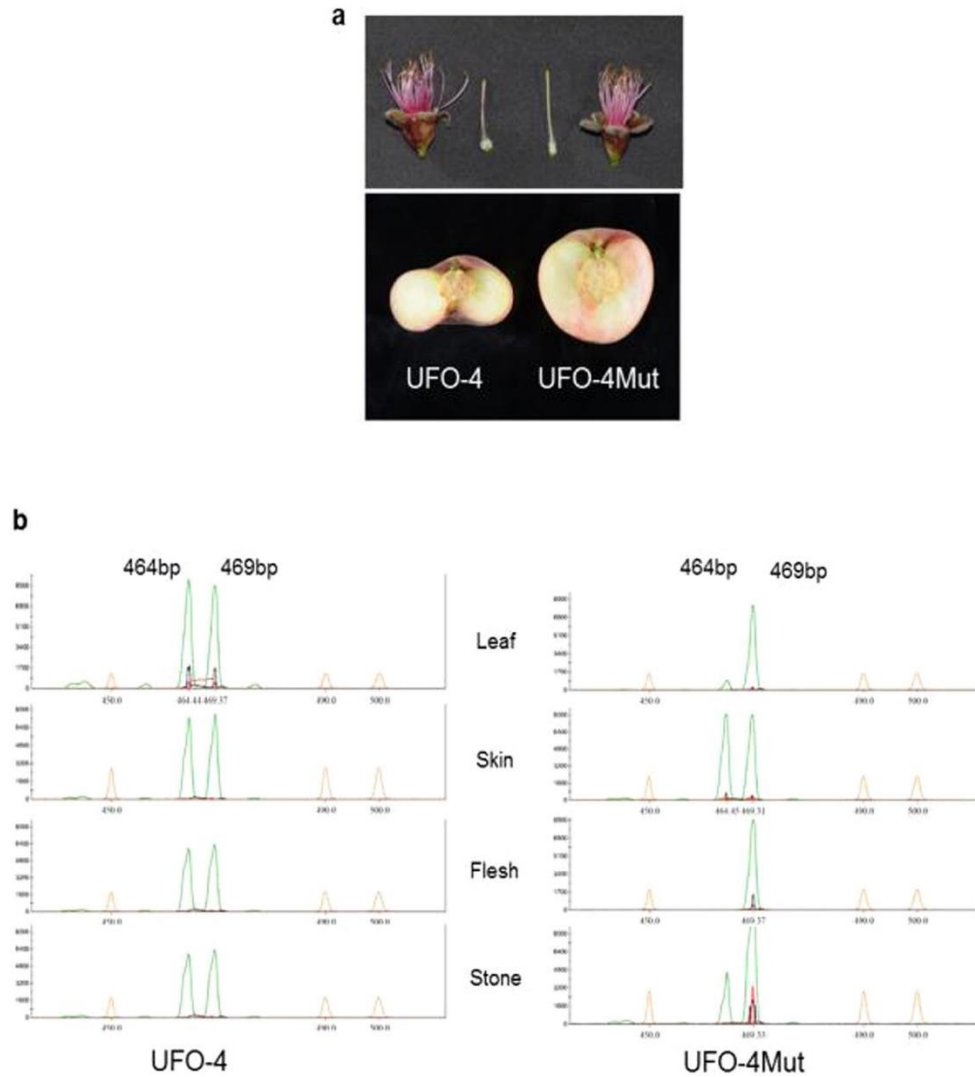
Figure 1.5. Analysis of a flat variety ('UFO-4MUT) and its somatic round mutant ('UFO-4Mut'). **a)** Image of the flat (left) and round (right) pistils and fruit. **b)** PCR-amplification products obtained with PC4 to detect the two small INDELs associated with the flat trait. PCR reactions were carried out with DNA extracted from leaf, as well as from skin, flesh and stone fruit tissues. The peaks show that the mutation occurred in flesh tissue (meristematic layer LII) affecting the flat associated allele.

Chapter 2: Characterization of genetic variability in LRR-RLKs clustering in the *S* locus

2.1 Background

In previous chapter 1 of this thesis (López-Girona *et al.*, 2017) we proposed the gene *Prupe.6G281100* as candidate for peach fruit flat shape. Briefly, we identified a 10 Kb deletion affecting the promoter and part of the CDS of this gene co-segregating in heterozygosis with flat shape, while the deletion in homozygosis was always observed in aborting fruit. Gene expression analysis and the analysis of a mutant supported the candidate gene hypothesis. To validate the deletion in germplasm, we designed a molecular marker yielding PCR fragments easily scored in agarose gels. This molecular marker is already used in parental lines and progenies of peach breeding programs for marker assisted selection (MAS). The routine use of this molecular marker in germplasm has identified three cultivars escaping from the association. One of the cultivars, a parental line in a breeding program encoded as M14, contains the 10 Kb deletion in heterozygosis as observed in all flat peaches, but shows a shape intermediate between flat and round. A second cultivar with unexpected phenotype is M247, first generation offspring from M14, with flat viable fruit despite containing the deletion in homozygosis. However, the fruit set and production of M247 is much reduced in comparison with other flat peach commercial varieties. The third cultivar escaping the association is 'Cerrito', a Brazilian variety characterized by its low chilling requirement with round shape in warm regions although with tendency to became oblate in colder conditions. 'Cerrito' shows the 10 Kb deletion in heterozygosis. Pictures of M14, M247 and 'Cerrito' are shown in Figure 2.1.

Figure 2.1 The front and lateral view of peach varieties. M14 (left up), its flat offspring M247 (left down), and 'Cerrito' (right).

Recently Cao *et al.* (Cao *et al.*, 2016) have reported some Chinese varieties escaping also from this association, and suggesting another gene 650 Kb downstream *Prupe.6G281100* that may act alone or in combination with other genes as an independent cause of the flat shape. Therefore, these inconsistences pinpoint the need

of exploring the variabilities in the *S* locus to understand the reasons for the unexpected phenotypes. With this aim in this chapter we have re-sequenced with Illumina technology the whole genome of three cultivars and explored the DNA variabilities within the *S* locus, and particularly around *Prupe.6G281100.*

2.2 Materials and methods

2.2.1 Plant material for DNA and RNA isolation

Genomic DNA from peach samples M14, M247, 'Cerrito', 'Red Diamond' (round), 'Mesmebrine' (flat), 'P07F202A056' (aborting progeny from an open pollination of 'ASF08.81'), as well as a panel of 94 individuals was isolated from young leaves using a modified Doyle's method (Doyle, 1990). All DNA isolation was used for PCR amplification. In addition, leaf samples from M14, M247, 'Cerrito' and 'P07F202A056' were processed for DNA extraction with DNeasy® Plant Mini Kit (QIAGEN, Hilden, Germany) and sent for Illunima resequencing.

For RNA extraction, branches with flower buds and open flowers from round ('Red diamond'), flat ('Mesembrine') and aborting ('P07F202A056') peaches were cut in the field and carried to laboratory in fresh and humid conditions, where close flower buds and pistils detached from open flowers were frozen in liquid nitrogen and stored at -80°C. Leaves and slices of flesh from M14 and M247 fruit (collected and sent by PLANASA, Huelva) and from 'Cerrito' (collected and sent by Jesús García Brunton, IMIDA-Murcia) were frozen in liquid nitrogen and stored at -80 °C. Total RNA was isolated using the RNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's specification. DNA contamination was removed following TURBO DNA-free™ (Ambion®) kit, followed by quantification with NanoDrop® Spectrophotometer ND-1000. DNA integrity was visualized by 1% agarose gel electrophoresis. Information of all the plant materials used can be found in Table 2.1.

Table 2.1 Plant materials used for PCR, Illumina re-sequencing and RT-PCR.

Sample	DNA			RNA			Source of materials
	Leaf (for PCR amplification)	Leaf (for Illumina resequencing)	leaf	Fruit flesh	Flower bud	Pistil	
M14	x	x	x	x			PLANASA, Huelva
M247	x	x	x	x			PLANASA, Huelva
Cerrito	x	x	x	x			IMIDA, Murcia
Red Diamond	x			x	x	x	IRTA, Gimenells
Mesembrine	x			x	x	x	IRTA, Gimenells
P07F202A056	x	x	x		x	x	IRTA, Gimenells
94 individuals plate	x						PLANASA, Huelva

2.2.2 Next-Generation sequencing and variability analysis

Genomic DNA of M247, M14, 'Cerrito' and 'P07F202A056' was delivered to CNAG (Centre Nacional d'Anàlisis Genòmica, Barcelona) for library preparation and 2 x 150 bp paired-end sequencing with Illumina HiSeq.3000/4000 sequencer. Adapter removal, quality trimming and mapping to the peach genome version 2.0 was done as in chapter 1. Whole genome alignment files, together with peach genome v2.0, annotated genes and transcripts files downloaded from Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html; Goodstein et al. 2012), were imported into *CLC Genomics workbench v10.1.1* for coverage and variability analysis using Built-in tools, as well as for track visualization.

2.2.3 Validation of *in silico* polymorphisms identified in *Prupe.6G281400* and *Prupe.6G281500*

2.2.3.1 Primer design and use

To validate the polymorphisms identified *in-silico* affecting *Prupe.6G281400* (Pp06:26281266..26290418) and *Prupe.6G281500* (Pp06:26292741..26296704), we designed primers using web tool *Primer3* (https://www.rosaceae.org/tools/primer3) to amplify by PCR the corresponding regions and sequence them by Sanger Primer information can be found in Table 2.2, Supplementary Table 2.S1 and Supplementary Figure 2.S1. For *Prupe.6G281500*, we designed the primer pair Prupe15_F/Prupe15_R, with both primers flanking a putative deletion of 1 Kb within the gene. The fragment size expected according to the reference genome was of 3.7 Kb. Both, forward and reverse primers, were used for PCR amplification and Sanger sequencing. For *Prupe.6G281400*

we designed the primer pair Prupe14_1F/Prupe14_1R, with both primers flanking a putative deletion of close to 6.1 Kb affecting part of the 5' region of the gene. The fragment size expected according to the reference genome was of around 9 Kb. Besides, primers Prupe14_1F and Prupe14_1R together with internal ones to the amplified regions (Prupe14_2F, Prupe14_3F, Prupe14_4F, Prupe14_5F and Prupe14_NDF) were used for Sanger sequencing (Table 2.2).

To explore the variability upstream and downstream *Prupe.6G281400* we used 4 primer combination yielding large sized products as well as internal primers for sequencing. Upstream variability was evaluated with the two primer-pair combinations: Prupe14_1F/Prupe14_3R (expected bands of 8.6 Kb in the reference genome and 2.5 Kb in the varieties with the deletion) and Prupe14_5F/ Prupe14_2R (yielding bands of 10.1 Kb in the reference gnome and 4 Kb when carrying the deletion). Variability downstream *Prupe.6G281400* was evaluated with Primer pair Prupe14_RTF + Prupe14_2R (yielding a band of 2.9 Kb). To preferentially amplify the shorter band (i.e. the allele carrying the deletion) we used limited extension time. In order to distinguish polymorphisms in phase with the deletion in the gene *Prupe.6G281100*, we used the primer combinations Prupe14_1F/Prupe14_NR (yielding a fragment of 8.4 Kb) and Prupe14_2F/Prupe14_dwR (yielding a fragment of 11.9 Kb) in the aborting genotype. Amplified and purified fragments were sent for sequencing with primers internal to the upstream fragments (Prupe14_1F, Prupe14_3F, Prupe14_4F, Prupe14_5F, Prupe14_NDF) and with primers internal to the downstream fragments (Prupe14_3R, Prupe14_eR, Prupe14_dR, Prupe14_cR, Prupe14_bR, Prupe14_aR and Prupe14_2R) (Table 2.2).

Table 2.2 Sequencing primers of amplicons to be genotyped for variabilities discovery.

Peach variety	PCR amplicon	Sequencing primers							LRR-RLKs validated
M14	Prupe15_F + Prupe15_R	Prupe 15_F	Prupe 15_R						*Prupe.6G281500*
M14	Prupe14_1F + Prupe14_3R	Prupe 14_1F	Prupe 14_3F	Prupe 14_4F	Prupe 14_5F	Prupe14 _NDF			
Aborting	Prupe14_1F + Prupe14_NR								
Aborting	Prupe14_2F + Prupe14_dwR								*Prupe.6G281400*
M14	Prupe14_5F + Prupe14_2R	Prupe 14_3R	Prupe 14_eR	Prupe 14_dR	Prupe 14_cR	Prupe14 _bR	Prupe 14_aR	Prupe 14_2R	
M14	Prupe14_RTF + Prupe14_2R								

2.2.3.2 PCR amplification

Long range amplifications we performed in a total volume of 25 µL with 40 ng genomic DNA, 1X LongAmp buffer, 300 µM dNTPs, 0.4 µM of each primer and 2.5 U *Taq* polymerase (New England BioLabs ®Inc.). The following reaction was performed on a S-1000™ Thermal Cycler (Bio-Rad Laboratories, Inc.; Hercules, California, USA): 94°C for 5 min; 35 cycles of 94°C (30 sec), 56°C/58°C (30 sec), 65°C (12 min); followed by a final extension at 65°C for 10 min. All PCR amplicons were checked on 1% agarose gel in TAE buffer and stained with ethidium bromide for band visualization.

For preferential amplification of the shorter allele, separate PCR reactions were set for each of the primer pair following protocol described previously with a few modifications: annealing temperature (57.5°C) and extension time (3min for Prupe14_1F + Prupe14_3R and Prupe14_RTF + Prupe14_2R; 5min for Prupe14_3F + Prupe14_2R; 9min for Prupe14_1F + Prupe14_NR and 11min for Prupe14_2F + Prupe14_dwR).

2.2.3.3 Sanger sequencing

PCR products yielding a single band were purified with the sepharose High Pure PCR product purification kit (Roche Diagnostic, Basel, Switzerland). For those PCR reactions yielding heterozygous results (i.e. two bands) we cut each of the bands separated in the agarose gel and purified them with the same kit. A total of 50 ng of purified PCR product were used as template for sequencing with 10 mM of primer and the BigDyeTMTerminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) using the following reaction: initial denaturation at 96 °C for 1 m, followed by 25 cycles of 96 °C for 10 s, 50 °C for 6 s, and 60 °C for 4 min. The sequences obtained with an ABI Prism 3130 × l DNA Analyzer (Applied Biosystems, Foster City, California, CA, USA) were visualized and manually edited with Sequencher 4.10.1 software (Gene Codes Corporation; Ann Arbor, MI, USA). The primers used for sequencing are described in Table 2.2.

2.2.4 Design of markers for genotyping

To genotype the 6.1 Kb deletion identified at *Prupe.6G281400*, we designed a three-primer combination, with two primers flanking the deletion (one forward Prupe14_NDF and one reverse Prupe14_NR, Figure 2.2) to obtain the allele containing the deletion (product size of 374 bp) and a third primer inside the deletion (Prupe14_NintF) to obtain, in combination with Prupe14_NR, a band of 194 bp in absence of the deletion. A total 10 µL volume PCR reaction was set, containing 20 ng genomic DNA, 1X NH_4 Reaction buffer, 200 µM dNTPs, 1.5 mM $MgCl_2$, 0.2 µM of each primer and 1 U BIOTAQ (Bioline). The following reaction was performed on a S-1000™ Thermal Cycler (Bio-Rad Laboratories, Inc.; Hercules, California, USA): 94°C for 2 min; 35

cycles of 94°C (20 sec), 57°C (20 sec), 72°C (20 sec); followed by a final extension at 72°C for 5 min. All PCR amplicons were checked on 1.8% agarose gel in TAE buffer and stained with ethidium bromide for band visualization.
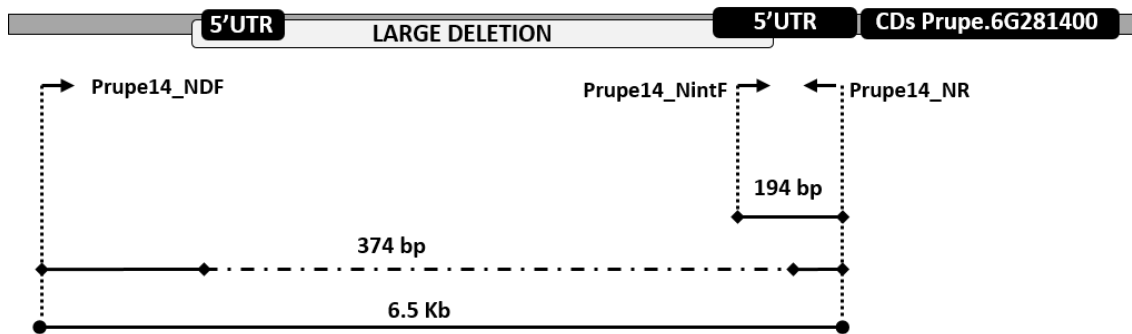
Figure 2.2 Schematic view of nested-PCR for genotyping the deletion. Grey bar represents the allele with or without the large deletion. Black boxes are *Prupe.6G281400* transcripts, including 5'UTRs and CDs. Arrows indicate primers position, while slid lines indicate fragments amplified by nested primers.

2.2.5 Transcription analysis of *Prupe.6G281400*

The Prupe14_sRTR primer designed in the CDS of *Prupe.6G281400*, 1,052 bp upstream the START codon, was used to reverse-transcript RNA to cDNA using the following protocol in a total 13 µL reaction volume: 1 µL of RNA was hybridized with 2 µL of 10 mM primer. After 10 min incubation at 70°C and 5 min cooling on ice, cDNA was obtained using PrimeScript RT-PCR kit (Takara).

Primer pair Prupe14_sRTF/Prupe14_sRTR was used to amplify aforementioned cDNA products with an expected size of 134 bp. This fragment contained SNPs in linkage with the deletion. The PCR reaction was conducted in a total volume of 10 µL, containing 20 ng of the correspondent cDNA, 1X NH$_4$ Reaction buffer, 200 µM dNTPs, 1.5 mM MgCl$_2$, 0.2 µM of each primer and 1 U BIOTAQ (Bioline). The following reaction was performed on a S-1000^{TM} Thermal Cycler: 94°C for 2 min; 35 cycles of 94°C (15 sec), 57°C (15 sec), 72°C (7 sec); followed by a final extension at 72°C for 5 min. All PCR amplicons were checked on 1.8% agarose gel in TAE buffer and stained with ethidium bromide for band visualization. PCR purification and sanger sequencing were performed as previously described. Sequencing results were trimmed and aligned to the reference genome as well as to allelic sequences obtained for further SNPs identification.

2.2.6 Gene functional prediction and protein sequence analysis

Functional   annotation   and   orthologues   for   *Prupe.6G281400*   and *Prupe.6G281500*   were   found   with   Dicots   PLAZA   3.0 (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/   (Proost   *et   al.,*

2015); Basic local alignment (blastp program) against PLAZA protein sequences database was carried out for similarity searches by using the BLOSUM62 score matrix (Altschul *et al.*, 1997; Schäffer *et al.*, 2001). DNA sequence of CDS of the LRR-RLKs was *in silico* translated in ExPASy nucleotide/protein sequence translation tool (http://web.expasy.org/translate/). Similarity analysis was performed against the non-redundant protein sequences database (https://blast.ncbi.nlm.nih.gov/Blast.cgi), using the blastp 2.8.0+ program (Altschul *et al.*, 1997; Altschul *et al.*, 2005). The algorithm parameters used to evaluate the quality of the pairwise sequence alignment was set with a BLOSUM62 protein substitution matrix allowing gap costs of 11 and an extension value of 1 under a conditional compositional score matrix adjustment.

Pairwise sequence alignment of the *Prupe.6G281400* CDS containing the deletion and the reference allele (missing the deletion) was performed with an online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) by setting the default parameters. Protein sequence analysis and classification for both two alleles was analyzed using InterPro online tool (http://www.ebi.ac.uk/interpro/) where the signature matches for domains and repeats of the proteins were generated.

2.3 Results

2.3.1 Analysis of polymorphism in the *S* locus

2.3.1.1 Whole genome re-sequencing

We re-sequenced with Illumina technology the three varieties with phenotype in discrepancy with their genotype in *Prupe.6G281100* (M14, M247 and 'Cerrito') as well as the aborting genotype 'P07F202A056'. All the reads were trimmed and filtered to remove low quality sequences, yielding an average of 14,377 Gb per sample, which represents a coverage of 57.51 times the peach genome. High quality of reads were aligned to the peach reference genome v2.1 constructed from the sequence of the peach variety 'Lovell' (Verde *et al.*, 2013). Sequences alignment analysis revealed a low percentage of singletons (i.e. pairs of reads with only one of the mate reads mapping), while the majority of mate reads mapped correctly in pairs, which should allow the detection of the different types of small structural variations (SVs).

The visualization of the sequence alignments with CLC software confirmed the 10 Kb deletion affecting *Prupe.6G281100* in heterozygosis in M14 and in 'Cerrito' and in homozygosis in M247. In addition, this visualization allowed detecting additional INDELs in the genes *Prupe.6G281400* and *Prupe.6G281500*, both LRR-RLK located 15.3 Kb and 21.3 Kb downstream *Prupe.6G28110*, respectively (Figure 2.3). In the case of the first gene, we observed in M14, M247 and 'Cerrito' a reduction of reads of about 6.1 Kb, suggesting a deletion in heterozygosis. This deletion was not evident in the sequences

of round, flat or aborting control samples. In the case of the second gene, the absence (in M247) or reduction (in M14 and 'Cerrito') of reads in a 1 Kb region suggested a deletion in homozygosis in M247 or in heterozygosis in other two samples. This reduction of reads was also observed in the alignments of flat and aborting samples, suggesting that this polymorphism was associated with the *S* allele.
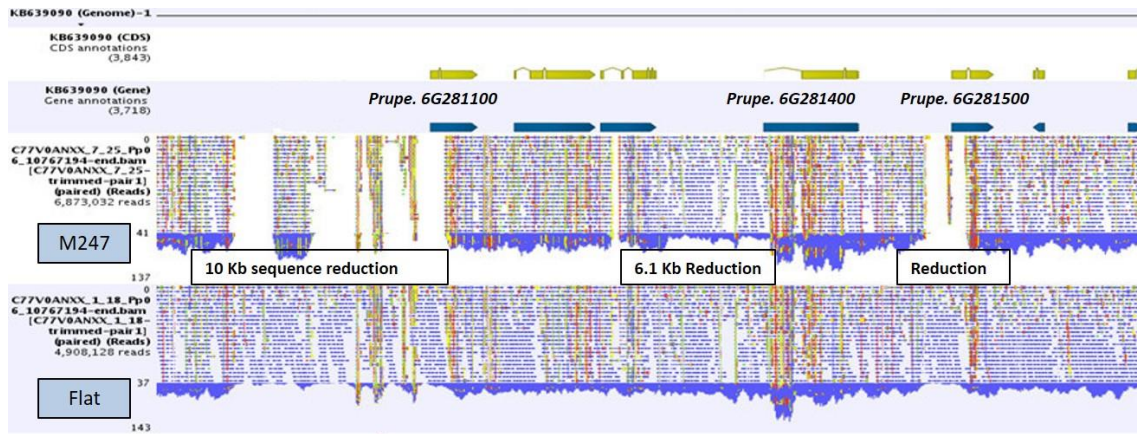
Figure 2.3 Illumina resequencing data analyzed in CLC-workbench showing overview for peach varieties (Flat as a control) LRR-RL Kinases region aligned to reference genome. Three regions with reads reduction affecting *Prupe.6G281100*, *Prupe.6G281400*, and *Prupe.6G281500* are indicated in box.

To validate the two putative deletions observed *in silico*, we amplified by PCR the affected regions. For this, a long-range PCR reaction with primers flanking the deletion in *Prupe.6G281400* (Prupe14_1F/Prupe14_1R) yielded the band of 9 Kb expected from the reference genome in round, flat and aborting samples, while amplified a shorter band (of approximately 3 Kb) in heterozygosis in M14, M247 and 'Cerrito' (Figure 2.4). On the contrary, the primer pair flanking the putative deletion affecting *Prupe.6G281500* gene (Prupe15_F /Prupe15_R) amplified in all samples bands of 3.7 Kb, as expected from the reference genome. To further analyze the amplified fragments in this later gene, we sequenced them with Sanger technology. The sequences revealed a region containing large number of polymorphisms in homozygosis in M247 and in the aborting sample, and in heterozygosis in M14, 'Cerrito' and flat varieties. The BLAST alignment of the sequence obtained in M247 and the aborting peach against the *Prunus persica* reference genome revealed a high degree of homology to *Prupe.6G281500* and *Prupe.6G281400*. Reads depth analysis in M247 and in the aborting revealed an increase of reads and of heterozygosity in *Prupe.6G281400*. Therefore, we must consider the reduction an increase in the reads depth in *Prupe.6G281500* and *Prupe.6G281400*, respectively, as the consequence of the wrong alignment of the allele in linkage with the deletion in *Prupe.6G281100*, and consequently, highly associated with the flat shape.
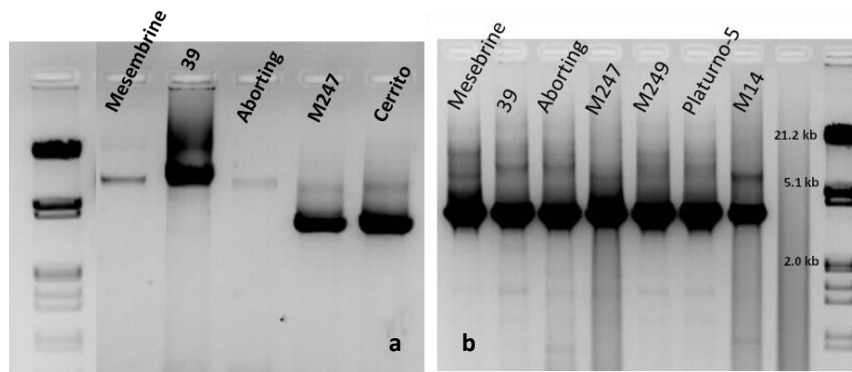
Figure 2.4 Long-Range PCR to validate the deletions affecting *Prupe.6G281500* and *Prupe.6G281400*. **a)** primers Prupe14_1F + Prupe14_1R with 10 µL loaded of PCR products (samples round (39), M247 and M14; Prupe14_1F + Prupe14_1R with 25µL loaded of PCR products (samples Flat (Mesembrine), Round MB39 (39), Aborting as controls, M247, and 'Cerrito'). **b)** 10 µL PCR products loading amplified with primers Prupe15_F + Prupe15_R (samples Flat (Mesembrine), Round MB39 (39), Aborting, Platurno-5, M249 as controls, M14, and M247). Digested lambda phage was used as ladder.

2.3.1.2 Analysis of the deletion affecting *Prupe.6G281400*

We sequenced the shorter band (i.e. the band containing a deletion) amplified with Prupe14_F and Prupe14_R and aligned it against the *Prunus persica* v2.0.a1 reference genome. The alignment indicated that the deletion affected the region Pp06:26,281,074..26,287,180 (6,106 bp long), starting 192 bp before the 5'UTR of the primary transcript and ending at the end of the 5'UTR, just 144 bp before the start codon of the gene. Therefore, this deletion covers most of the 5'UTR region, and perhaps the promoter of LRR-kinase *Prupe.6G281400*, without disturbing the coding part.

Within CLC, the analysis of variants along the region flanking *Prupe.6G281400* identified a high rate of polymorphisms up and downstream it. For the upstream region, all SNPs observed in M247 in homozygosis were also present in homozygosis in the aborting sample and in heterozygosis in the flat, M14 and 'Cerrito' DNA, therefore occurred in the flat allele background (i.e. in an allele containing the 10 Kb deletion and the polymorphisms in *Prupe.6G281100*). The analysis did not identify SNPs in heterozygosis in M247 (i.e. in linkage with the allele carrying the deletion). Some of these polymorphisms were validated by sequencing 5 fragments upstream the deletion (2.2 Kb, 1.6 Kb, 1.1 Kb, 0.4 Kb and 0.2 Kb upstream; average size 400 bp) in the aborting, round, flat, M247, M14 and 'Cerrito', showing a SNP frequency rate of 3% (3 SNPs per 100 bp). This region also included part of *Prupe.6G281300* (Figure 2.5).

CLC variants analysis showed much higher variability ratio in the region downstream the deletion, corresponding to the CDS region of the gene. As deduced from the analysis of *Prupe.6G281500*, most of this variability was due to a misalignment. Therefore, to infer the putatively real SNPs we filtered out those variants homozygous

in M247 and in the aborting sample (as both samples were homozygous for the misalignment in *Prupe.6G281500*). This strategy allowed to identify SNPs in M247 not present in the aborting allele; all were also observed in heterozygosis in M14 and 'Cerrito', and therefore, in linkage with the allele carrying the deletion. Some of these polymorphisms were validated by sequencing 7 fragments (0.3 Kb, 1.1 Kb, 1.6 Kb, 2.2 Kb, 2.7 Kb, 3.1 Kb and 3.6 Kb downstream the deletion; average size of 400 bp). Sanger sequencing confirmed the high SNP rates (16%) exclusive to the allele carrying the deletion (Figure 2.5).
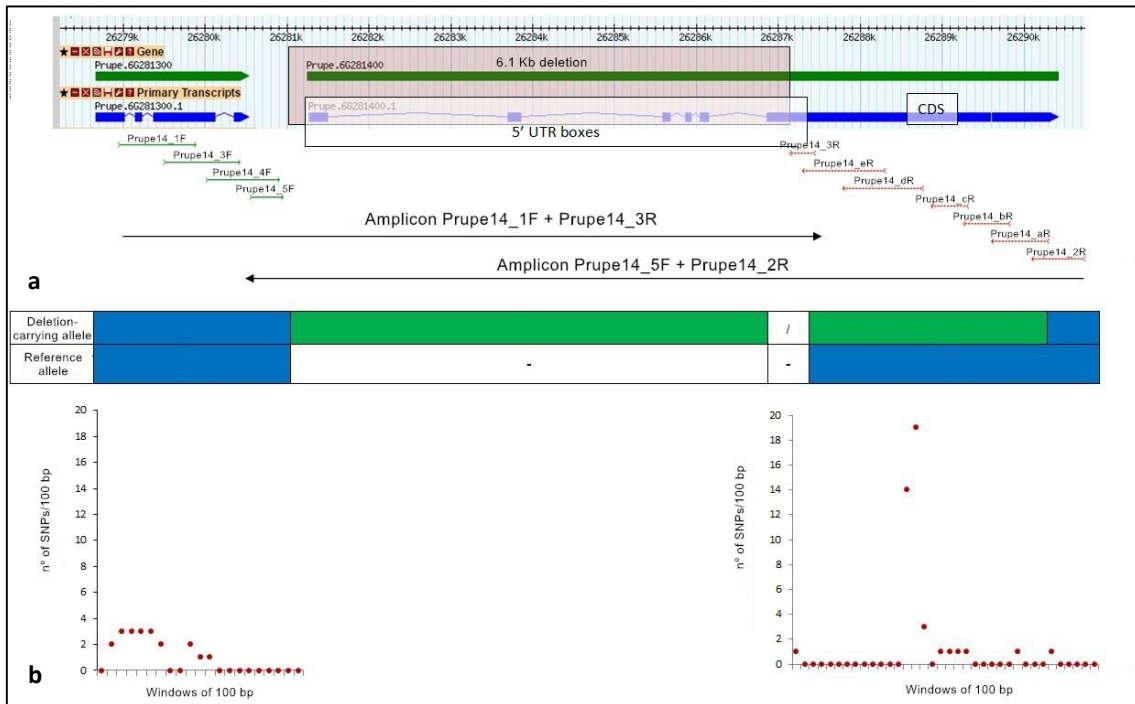
Figure 2.5 Genetic variability surrounding the big deletion regions (up to 2.2 Kb upstream and 3.5 Kb downstream) within the LRR-RLK cluster. **a)** Schematic pattern of the large deletion region affecting *Prupe.6G281400*. Green boxes correspond to genes, blue boxes are mature mRNA, interspersed blue lines are introns, 5'UTRs regions and parts of CDS. **b)** Color boxes combination represent haplotypes for the reference allele from aborting sample and for deletion-carrying allele from M14 sample. Blue boxes are identical to the aborting sequences, green boxes are different from reference or aborting sequences, '-' means lack of sequences, '/' means insufficient information to distinguish both alleles. Below scatter diagrams are SNPs calculated in 100 bp windows found in deletion-carrying allele in the regions surrounding the deletion (refer to Querol's master thesis).

To efficiently genotype the big deletion in larger sets of individuals we performed nested-PCR assay with two flanking and one internal primer to yield band of 194 bp for the reference allele and of 374 bp for the one containing the deletion. This marker was used in a collection of 94 accessions including M247 and M14 as controls. None of them (except for the controls) contained the 374 bp band, indicating that this deletion is uncommon in modern peach cultivars (Figure 2.6).
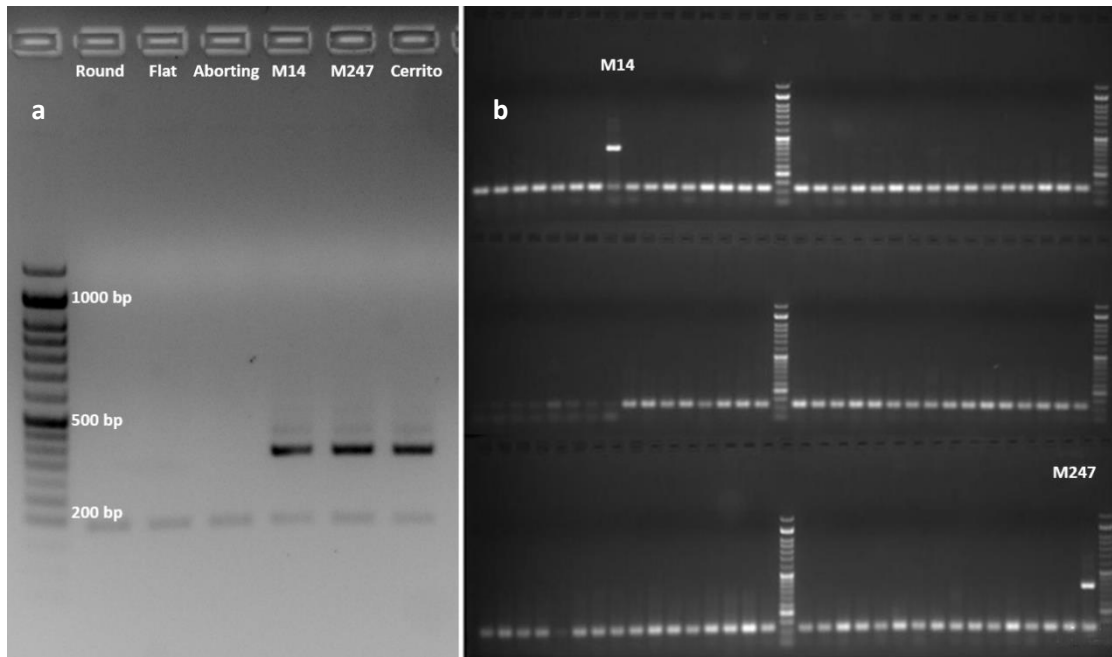
Figure 2.6 Nested-PCR for genotyping a big deletion in different peach varieties. **a)** PCR with PC1 (Prupe14_NDF/Prupe14_NintF/Prupe14_NR) yielding a short band of 194 bp in homozygosis in round, flat, aborting samples and in heterozygosis with a band of 374 bp in M14, M214 and 'Cerrito'. **b)** Screening of peach cultivars with PC1, M14 and M247 were used as controls.

2.3.2 *Prupe.6G281400* expression in various individuals and tissues

To verify if *Prupe.6G281400* expresses at the initial phases of flower development, when the fruit shape is already determined, we carried out reverse-transcription PCR using RNA from various tissues (leaves, flower buds, ovaries and fruit). As we identified several SNPs in heterozygosity in the coding regions of *Prupe.6G281400* unique to M14, M247 and 'Cerrito', we designed RT-PCR primers flanking a few of them to amplify a fragment including such SNPs (Table 2.3). A band of the expected size was present in all samples, including the positive control using genomic DNA (Figure 2.7). Sanger sequencing of the bands revealed that only the allele carrying the variants as in the reference genome was amplified in M14, M247 and in 'Cerrito' samples, indicating the lack of expression of the allele carrying the SNPs in linkage with the deletion. Therefore, these results suggest that the deletion in *Prupe.6G281400* prevented from gene expression.

Table 2.3 SNPs of *Prupe.6G281400* gene transcription from the reference allele and the deletion-carrying allele in either gDNA or cDNA of samples round, M14, M247 and 'Cerrito'. Nucleotides A, T, C, G are colored separately, while SNPs between the reference allele and deletion-carrying allele are highlight in yellow (refer to Querol's master thesis).

Prupe.6G281400 reference allele	C	A	T	T	:	C	A	A	T	C	T	T	C	T	G	A	C
Deletion-carrying allele	C	:	T	T	A	C	A	A	T	C	A	G	C	T	A	A	C
Round genomic	C	A	T	T	:	C	A	A	T	C	T	T	C	T	G	A	C
M14 genomic	C	A/:	T	T	:/A	C	A	A	T	C	T/A	T/G	C	T	G/A	A	C
M247 genomic	C	A/:	T	T	:/A	C	A	A	T	C	T/A	T/G	C	T	G/A	A	C
'Cerrito' genomic	C	A/:	T	T	:/A	C	A	A	T	C	T/A	T/G	C	T	G/A	A	C
Round cDNA	C	A	T	T	:	C	A	A	T	C	T	T	C	T	G	A	C
M14 cDNA	C	A	T	T	:	C	A	A	T	C	T	T	C	T	G	A	C
M247 cDNA	C	A	T	T	:	C	A	A	T	C	T	T	C	T	G	A	C
'Cerrito' cDNA	C	A	T	T	:	C	A	A	T	C	T	T	C	T	G	A	C

Figure 2.7 An initial analysis of gene expression for *Prupe.6G281400*. Agarose gel visualization shows a RT-PCR amplification band within different individuals, tissues during flower (left) and fruit development stage (right). Samples round (Ro), flat (Fl), aborting (Ab), M14 and M247 were collected in 2016 (leaves for gDNA control) and 2017 (others): flower bud in February, ovaries in March, immature fruit flesh in April and leaves in February, March, April and May (refer to Querol's master thesis).

2.3.3 Functional prediction and protein sequence analysis

2.3.3.1 Homology and functional prediction of the two LRR-RLKs

Dicots PLAZA 3.0 identifies *Prupe.6G281400* as member of the HOM03D000009 gene family, which has 3332 genes in 29 species, and a member of the ORTHO03D028407 subfamily, which has 4 genes in three species: one in apple, one in eucalyptus and two in peach (*Prupe.6G281400* and *Prupe.6G28880*, 407 Kb apart). It is described as receptor-like kinase protein *THICK TASSEL DWARF1*. Translated protein BLAST against the PLAZA protein sequence database produced best alignments with 250 genes of its family (HOM03D000009) of 24 species (Supplementary Table 2.S2), including

Prunus persica, Malus domestica, Fragaria vesca and *Citrus x sinensis*. In peach, those highly homologous transcripts are in chromosome 3 (*Prupe.3G094800*), chromosome 6 (*Prupe.6G281000*, *Prupe.6G281100*, *Prupe.6G281200*, *Prupe.6G281500*, *Prupe.6G288800*), chromosome 7 *(Prupe.7G088700)*, and chromosome 8 (*Prupe.8G054400, Prupe.8G054300)*. The top three orthologous with most gene blasts are ORTHO03D000261 (88), ORTHO03D000539 (79), ORTHO03D001987 (26), among which 96 (49.7%) were described as receptor-like kinase protein *THICK TASSEL DWARF1*, 34 as *BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1* (17.6%), 41 (21.2%) are leucine-rich repeat receptor-like kinase protein *FLORAL ORGAN NUMBER1*. The rest few annotations are repeat receptor protein kinase *EXS*, *Receptor-like protein kinase BRI1-like 3*, S-cell enriched with leucine-rich repeat-containing protein *slrA*.

The other LRR-RLK gene (*Prupe.6G281500)* was also categorized into HOM03D000009 gene family but within a different subfamily (ORTHO03D000261) which include also the peach genes in chromosome 7 an 8 *Prupe.7G088700, Prupe.8G054300 and Prupe.8G054400*. BLAST protein sequence in PLASA interface identified optimal alignment with 250 genes of twenty-six species, including the previous 10 peach homologous transcripts plus additional *Prupe.4G234200* in chromosome 4 (Supplementary Table 2.S3). Altogether, the orthologous ORTHO03D000261 with 85 genes, ORTHO03D000539 with 79 genes and ORTHO03D001987 with 26 genes account for 76% of the total 250 aligned sequences. As with the previous gene, the majority of the orthologous genes belonged to three subfamilies: 95 genes with description of receptor-like kinase protein *THICK TASSEL DWARF1* occupies the most portion (50%), followed by 41 receptor-like kinase protein *FLORAL ORGAN NUMBER1* (21.6%), 33 *BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE* 1 (17.4%). Inactive leucine-rich repeat receptor-like protein kinase, receptor like protein 6, Receptor-like protein kinase *BRI1-like 3*, repeat receptor protein kinase *EX*S and S-cell enriched with leucine-rich repeat-containing protein *slrA* depict other annotations.

2.3.3.2 Functional protein alteration by *Prupe.6G281400* nucleotide polymorphisms

Pairwise sequence alignment between *Prupe.6G281400* reference allele and that characteristic of M14, M247 and 'Cerrito' (deletion-carrying allele which contains the 6.1 Kb deletion and SNP polymorphisms) indicates that the later encodes a polypeptide 614 shorter, with 409 amino acids instead of the expected 1,023 (Figure 2.8 a). Protein sequence analysis and classification gave a prediction for *Prupe.6G281400* having 4 Leucine-rich repeat (LRR) homologous superfamily, 1 LRR-containing N-terminal domain, 11 Leucine-rich repeat motifs as well as some signal peptides. A region of a membrane-bound protein predicted to be embedded in the membrane was also found close to the N-terminal. Near the C- terminal, there was a region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region (Figure 2.8 b, Supplementary Table 2.S4). By contrast, the polypeptide 409 AA long lacked half of the

homologous superfamily, most of LRR motifs and domains, the transmembrane domain and the uncertain C-t regions. Therefore, in case of transcription, it is very unlikely to be functional.
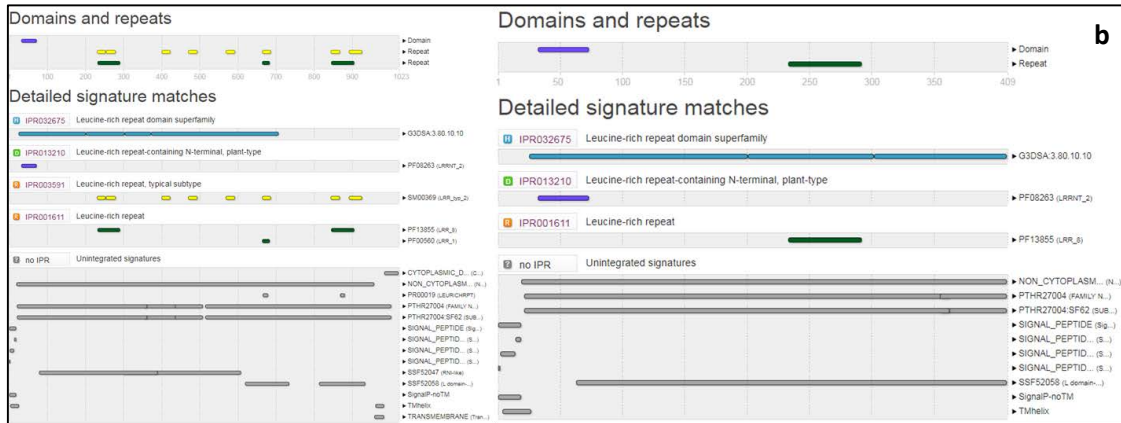
Figure 2.8 Predicted effect of deletion-carrying allele on Prupe.6G281400 amino acidic sequence and on functional domains and motifs. **a)** Protein sequence alignment of Prupe.6G281400 translated reference allele and deletion-carrying allele obtained by Sanger sequencing. Symbols represent: '*' match, ':' mismatch, ' ' mismatch of low biological relevance (AAs from the same family) and '-' one of the sequences is missing. **b)** Patterns of homologous superfamily, repeat motif, domain and peptide for Prupe.6G281400 (left) reference and deletion-carrying allele (right): homologous superfamily in navy, LRR-containing N-terminal domain in purple, LRR repeat motif in yellow and dark green, transmembrane domain and signal peptide in grey.

Chapter 3: Genetic characterization of a flat variety and its sport mutant which exhibits different fruit shape

3.1 Background

In the first chapter of this thesis we report a mutant ('UFO-4Mut') derived from a flat fruit variety ('UFO-4') exhibiting fruit with round shape (Figure 3.1). DNA analysis revealed that the mutation occurred in meristematic L2. In this chapter we aim at studying such mutation as a way to (i) validate the candidate gene and (ii) studying somatic mutation mechanisms in peach. For this we have re-sequenced with Illumina technology genomic DNA of the original flat and of the mutant round peaches.

Data analysis suggests a chromosome rearrangement producing a reduction of heterozygosis in the modified region (which at its time contains the *S* locus). To visually validate this hypothesis, we initiated a protocol of fluorescent in situ hybridization (FISH), encountering several limitations for the analysis of chimeric tissues. Therefore, in this chapter we study in depth somatic variability and provide some tips for future development of a FISH technique to identify rearrangement in chimeric tissue in peach.

3.2 Materials and methods

3.2.1 Plant material collection and DNA extraction

High quality DNA from 'UFO-4' and 'UFO-4Mut' (sport mutation from 'UFO-4'UFO-4') was extracted from young leaves collected from plants grafted in Torre Marimon (Caldes de Monbui, IRTA) using a modified Doyle's method (Doyle, 1990; Porebski *et al.*, 1997). In total 100 mg of young and fresh leaves were ground to powder in a 1.5 mL Eppendorf using plastic stick in liquid nitrogen. The mashed tissues of each Eppendorf were re-suspended in 500 µL CTAB buffer (10 mM Tris-HCl, 1.4 M NaCl, 20 mM EDTA, 2% CTAB) and heated at 65°C for 30 min. Vortex slowly and add 500 µL of chloroform: isoamyl alcohol (24:1) to one sample. The suspension was centrifuged at 11,000 rpm at room temperature (RT) for 10 min. The aqueous layer was transferred to a new Eppendorf and DNA was precipitated by adding 350 µL of pre-chilled isopropanol and gentle mix. DNA was then precipitated by centrifuging at 10,000 rpm at room temperature during 5 min and washed with 200 µL of Doyle cleaning solution (EtOH 75%), followed by centrifuge at 14000 rpm at RT during 10 min. After removing the supernatant, the DNA pellet was finally dried by inverting the tubes over laboratory tissue in a hood for 30 min to 1 h and re-suspended in 1x Tris EDTA. To clean RNA contamination, add 1 µL RNAse (10µg/µl) and vortex slowly, followed by 37 ºC incubation for 1 h. DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) was used for DNA extraction from peach fruit flesh, fruit skin, and plant tender root following the manufacture's protocol. Finally, the DNA quality was measured and evaluated with NanoDrop® Spectrophotometer ND-1000 and 1% agarose gel electrophoresis.

DNA from 'UFO-4' and 'UFO-4Mut' peach fruit mesocarp was extracted as in López-Girona *et al.* (López-Girona *et al.*, 2017). Young branches of 'UFO-4' and 'UFO-4Mut' were collected in spring and introduced to *in vitro* growth by the group of Ramon Dolcet in IRTA-FruitCentre (Lleida). After 6 weeks Healthy *in vitro* plants were induced rooting and maintained *in vitro*.

Figure 3.1 The front and lateral view of peach varieties 'UFO-4' and its mutant 'UFO-4Mut'.

3.2.2 Next-generation sequencing data analysis

3.2.2.1 Peach genome resequencing

Five micrograms of two replicates of high quality DNA of 'UFO-4'and 'UFO-4Mut' were sent to CNAG (Centre Nacional de Análisis Genómico", Barcelona, Spain) for Illumina sequencing. To prepare genomic DNA libraries, each genomic DNA was fragmented by nebulization to yield < 800 bp double strand fragments. After that, DNA fragments, repaired with adding "A" tails to 3' terminals, were ligated with paired-end adapter oligonucleotides from Illumina. Then those products were purified and enriched by PCR amplification. Each library was quantified by Agilent 2100 Bioanalyzer (Agilent, Foster city, USA) and sequenced by CNAG using Illumina HiSeq2000 platform.

3.2.2.2 Heterozygosity variants detection

Whole genome sequencing reads in FASTQ format were imported to *CLC-genomics workbench v8.5.1* (CLC bio, Qiagen, Aarhus, Denmark) and trimmed using a minimum phred score of 20 to improve the quality of assembly bases at the end of reads. The trimmed reads of each two independent genomes were assembled using *CLC*'s *de novo* assembly algorithm, using an appropriate k-mer and a minimum scaffold length of 1 Kb (Garg *et al.*, 2011; Albertsen *et al.*, 2013). To map all the scaffolds to reference

genome, we set the following parameters in *CLC-genomics workbench*. Masking mode: no masking; mismatch cost: 2; insertion cost: 3; deletion cost: 3; length fraction: 0.5; similarity fraction: 0.9; global alignment: no; auto-detect paired distances: yes; non-specific match handling: map randomly (Hawkins *et al.*, 2016). Two built-in tools *InDels* and *Structural Variants* were run separately to evaluate heterozygous variants in the alignments using default parameters (unaligned end breakpoints, *P*-value threshold = 0.0001, Maximum number mismatches = 3).

3.2.3 Peach somatic mutation genotyping

Based on NGS data, heterozygosity within approximately 6.5 Mb of distal end region of chromosome 6 in 'UFO-4Mut' was missing. To confirm *in silico* data, we designed primers along the 6.5 Mb every 1 Mb (approximately) in regions with heterozygous polymorphisms in 'UFO-4Mut'. For this we designed 7 primer pairs with *Primer3web* tool (https://www.rosaceae.org/tools/primer3) to yield products from 505 bp to 778 bp that were used to validate SNPs (Supplementary Sequence 3.S1, Table 3.1).

To validate the starting point of the mutation we designed a primer pair (LOH_sv1F and LOH_sv1R) to amplify a 5 Kb region including such area. These two primers, together with primers inside the region LOH_sv3R and LOH_sv1F (Table 3.2) were used for Sanger sequencing.

A total 10 µL volume PCR reaction was set, containing 20 ng genomic DNA, 1x LongAmp buffer, 200 µM dNTPs, 1.5 mM Mg^{2+}, 0.2 µM of each primer and 1 U *Taq* polymerase (New England BioLabs ®Inc.). The following reaction was performed on a S-1000™ Thermal Cycler (Bio-Rad Laboratories, Inc.; Hercules, California, USA): 95°C for 2 min denaturation; 35 cycles of 94°C (15 sec), 55-60°C (15 sec), 72°C (30 s); followed by a final extension at 72°C for 7 min. All PCR amplicons were checked on 1.5% agarose gel in TAE buffer and stained with ethidium bromide for band visualization. PCR products were purified with the High Pure PCR product purification kit (Roche Diagnostic, Basel, Switzerland) directly or after cutting the corresponding agarose gel bands. Purified bands were sequenced with Sanger technology. The high-quality sequencing reads obtained were manually curated for quality check and aligned against the reference sequences in *Sequencher v4.10.1*.

Table 3.1 Primers used to confirm LOH in 'UFO-4Mut'.

Primer name	Position of 5' start (GDR v2.0.a1)	Sequence (5' -> 3')	SNPs in amplicon
LOH_Con2F	25256172	ACTGCCAACATGTCCTCCTC	G/T
LOH_Con2R	25256837	GCGTTTGCCTCCTCTATTTG	
LOH_Con3F	26227383	ACAGAATTCAGCCGATGGAA	C/T
LOH_Con3R	26228464	TGCAGGTCGCAAATAAATGA	
LOH_Con9F	26924276	TTCTCGAATATGGGCTCCTG	A/T
LOH_Con9R	26924266	TCGATTCTGCTTGCTTCCTT	
LOH_Con10F	27488016	GAGTCTTGGTCTGCATCTTGG	T/C
LOH_Con10R	27488553	TCATGACATATGAGGAAGTGGA	
LOH_Con11F	28013592	GCAGTGCAAGATGCAACAAC	A/T
LOH_Con11R	28014127	GGGCCTTTCTGTATCCAGTG	
LOH_Con6F	28558467	TGACCCGTCCTTTTCTATGG	T/C
LOH_Con6R	28558099	TGTTGAGGGCGATTAGATCC	
LOH_Con12F	29613605	TGCATCACAGCTCCAAAAAC	G/A
LOH_Con12R	29614188	CTCCTTCTCCGTTCCGTACA	

Table 3.2 Primers used to validate the start region of the mutation.

Primer name	Position of 5' start (GDR v2.0.a1)	Sequence (5' -> 3')	SNPs in amplicon
LOH_sv1F	25256172	ACTGCCAACATGTCCTCCTC	G/T
LOH_sv3R	25256837	GCGTTTGCCTCCTCTATTTG	
LOH_sv4F	26227383	ACAGAATTCAGCCGATGGAA	C/T
LOH_sv1R	26228464	TGCAGGTCGCAAATAAATGA	

3.2.4 FISH preparation

3.2.4.1 Chromosome preparation

Root tips, collected from *in vitro* 'UFO-4'and 'UFO-4Mut' plants, were used to obtain chromosomes in metaphase. For this, they were detached from the plants and pre-treated with a mix of 0.1% colchicine (final 0.1 µg/mL) and 2 mM 8-hydroxyquinoline during 4 hours, followed by a fixation treatment with Carnoy's solution (methanol:acetic acid = 3:1) during 30-50 min. Fixed tissues were conserved in ethanol for further treatment, which involved tissue mechanical disruption and enzymatic tissue dissociation using an enzyme mixture (1.2% Pectolyase Y-23 (Kikkoman, Tokyo, Japan), 1.2% Cellulase Onozuka R-10 (Yakult Co. Ltd., Tokyo, Japan) and 1.2% Cytohelicase (Sigma-Aldish Co.LLC, France) prepared in citric buffer. After 2 hours of enzyme treatment, a drop 18–22 µL of cell suspension was dropped onto a microscope slide till the surface become granule-like, and the layer of fixative becomes thin (25–35 sec) (Ma *et al.*, 1996; Kirov *et al.*, 2014). Dry slides were stained with DAPI (4',6-diamino-2-

fenilindol) and visualized with the epifluorescent upright microscope Axiophot Zeiss equipped with the digital camera DP70 Olympus. The fluorescent filter used allowed DAPI detection in the range of 395-440 nm.

3.2.4.2 Probes design to identify mitotic recombination in 'UFO-4Mut'

FISH probes were designed to label the 10 Kb deleted region affecting the candidate flat shape gene and to validate the hybridization method. For this we used the whole-genome sequence alignment of both original and mutated plants to design the appropriate probes. Probe 1 (5,188 bp) was designed within the 10 Kb deletion; Probe 2 and Probe 3 were designed in regions with polymorphic indels up-and downstream the 10 Kb deletion, respectively (Table 3.3). These probes should hybridize along the distal end of the chromosome 6 carrying the round allele.

Table 3.3 Primers used to confirm deletion and obtain probes.

Amplicons	Primers name	Position of 5' start (GDR v2.0.a1)	Sequence (5'-3')	Fragmetn size (bp)
	LOH_DelupF	26022772	CCGCTCTACCCTCTCTACCA	Around 3-4 k
	LOH_DeldwR	26039168	TGTCCTGCATGGGATACTTG	
Probe1	LOH_Pr1F	26027113	TCTTCCTTCAGAGGCTTCCA	5188
	LOH_Pr1R	26032281	CGTCCAGAAAACCAACCAGT	
Probe2	LOH_Pr2F	26011100	AAACCTTTCAGCACCCATTG	4904
	LOH_Pr2R	26015984	TCATGAGGGGAAAAGCAATC	
Probe3	LOH_Pr3F	26050003	ATGCGCTCAGCAAGTAAGGT	4803
	LOH_Pr3R	26054786	TCTGGTTCATCAATGCGTGT	

3.2.4.3 Probes labelling and *in situ* hybridization

Probes were obtained by PCR from peach UFO-4 cultivar. Long-Range PCR reaction was set as following: 94°C for 5 min; 35 cycles of 94°C (30 sec), 56°C (30 sec), 65°C (12 min); followed by a final extension at 65°C for 10 min. All PCR products were gel cut and purified for further labeling with fluorescence using Nick Translation labelling system following the manufacturer protocol (Sumner *et al.*, 1990). In this technique DNA polymerase is used to replace some nucleotides of the probe sequence to produce single-stranded tagged sequences for further fluorescence hybridization. All three probes were labeled with Cyanine 3- (Cy3) fluorescent dye, which is detected in the green spectrum channel.

Fluorescence labeled probes were added to the microscope slide containing the cells and covered with a coverslip without leaving any bubbles. Slides were incubated at 37°C in a humid chamber at least overnight (hybridization can be left up to 3 days) and washed for three times with 50% formamide for 5 minutes each at 42°C, and three additional times with 2x SSC for 5 minutes at 42 °C. Finally, a drop of DAPI was added to the dry slides for microscope observation.

3.3 Results

3.3.1 NGS heterozygote variants calling for the mutation

To identify genetic changes in the 'UFO-4'mutant ('UFO-4Mut') leading to a new phenotype we sequenced genomic DNA from leaves of the original and mutant genotypes. Leaf tissue in 'UFO-4Mut' is chimeric, it is to say, cells of LII tissue contain mutated chromosomes while cells of LI (and probably LIII) tissue cohabit in the leaves. To overcome the limitations caused by such DNA chimerism we sequenced two replicates of each sample, obtaining a total sequence yield after trimming of 23.9 Gb for 'UFO-4' and of 22.3 Gb for 'UFO-4Mut', which represented a sequence depth of 95.7x and 89.3x, respectively. Sequence reads were aligned against the reference genome and analyzed in *CLC-genomics workbench v8.5.1* for variability. Each sample sequence was first individually compared with the reference genome, identifying SNPs, indels and structural variants along all chromosomes. When comparing variability between 'UFO-4' and 'UFO-4Mut' we identified low number variants. Such variants were equally distributed in all chromosomes with the exception of a large region, of approximately 6.5 Mb, at the distal end of chromosome 6 where most of heterozygous variants between 'UFO-4' and the peach reference genome disappeared or where homozygous in 'UFO-4Mut' (Figure 3.2, Table 3.3). In total the variant calling pipeline identified 1,590 indels and 12,538 SNPs between 'UFO-4' and the peach reference genome in the region Pp:26,281,253..26,290,418. Among them, 83% and 87% of indels and SNPs, respectively, were heterozygous. When compared 'UFO-4Mut' sequence with the peach reference genome these figures were reduced to 927 indels and 3,112 SNPs (58% and 25% of the ones observed in the original variety). Only a small number of these polymorphisms (255 indels and 1042 SNPs, 19% and 10% of the indels and SNPs in heterozygosis in the original, respectively) where heterozygous. Therefore, the number of heterozygous loci in the mutant experienced an 89% reduction. By the contrary, the number of homozygous loci increased in the mutant. This region contains the *S* locus. Therefore, a region of 6.5 Mb at the distal end of chromosome 6 presented a loss of heterozygosity (LOH).

Table 3.4 Small indels and SNP polymorphisms in heterozygosis (He) in homozygosis (Ho) between 'UFO-4' and 'UFO-4Mut' and peach reference genome.

Samples	Indels		SNPs	
	H_e	H_o	H_e	H_o
UFO-4	1,314	276	10,895	1,633
UFO-4Mut	255	672	1,042	5,182

Two possible reasons may explain this LOH. One involves the deletion of the whole 6.5Mb regions, leading hemizygosity. To test this hypothesis, we analyzed the sequence depth in this region with CLC. Such analysis as well as track visualization showed similar number of sequence reads in both original and mutated samples, discarding this hypothesis. The second possible reason involves somatic recombination between homologous chromosomes during mitosis, or the deletion of a chromosome region repaired by recombining using the sister chromosome (Figure 3.3). The extreme levels of variation observed only at the distal end of the chromosome suggested the second option (deletion plus repair with homologous chromosome) as the most plausible. In that case, a possible DSB affecting the chromosome carrying the flat allele may have been repaired by recombining with the homologous chromosome carrying the round allele, which will explain the new phenotype.
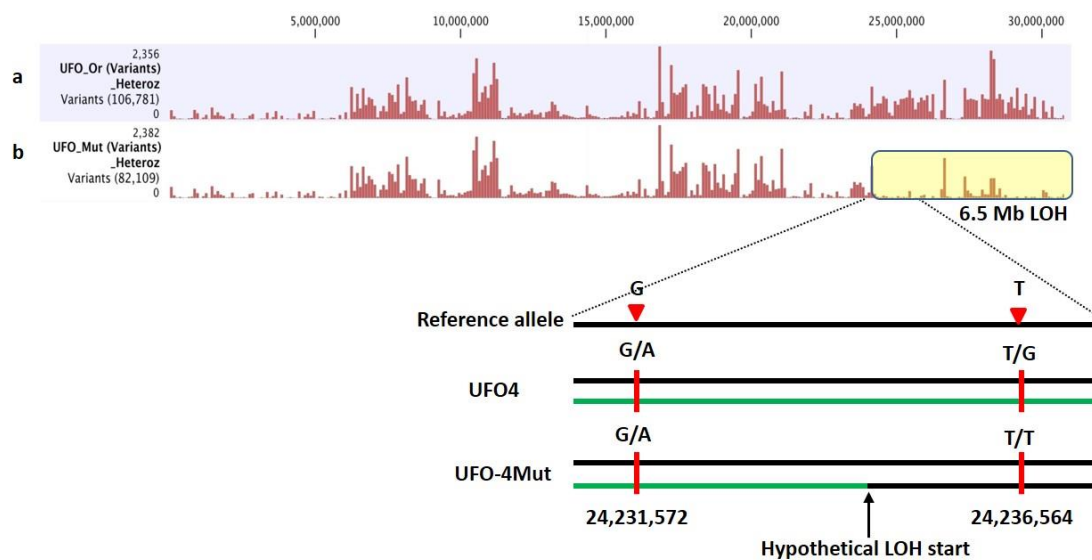
Figure 3.2 Distribution of heterozygous loci along chromosome 6 in 'UFO-4' **a)** and 'UFO-4Mut' **b)** samples. Red bars represent heterozygous polymorphisms; the heights of the bars are proportional to the number of polymorphisms accumulated at each position. This figure displays a loss of heterozygosis (LOH) at the distal end of the chromosome in the mutant sample, affecting a region of 6.5 Mb. Alleles at the starting point of the chromosome change, in phase with the round and flat alleles, are indicated in black and green bold lines in which red triangles and red circles depict SNPs alleles respectively. SNP G/A at Pp06:24,231,572 and SNP T/G at Pp06:24,236,564 are the closest heterozygous and homozygous loci detected flanking the hypothetical LOH starting point in 'UFO-4Mut' based on NGS data.
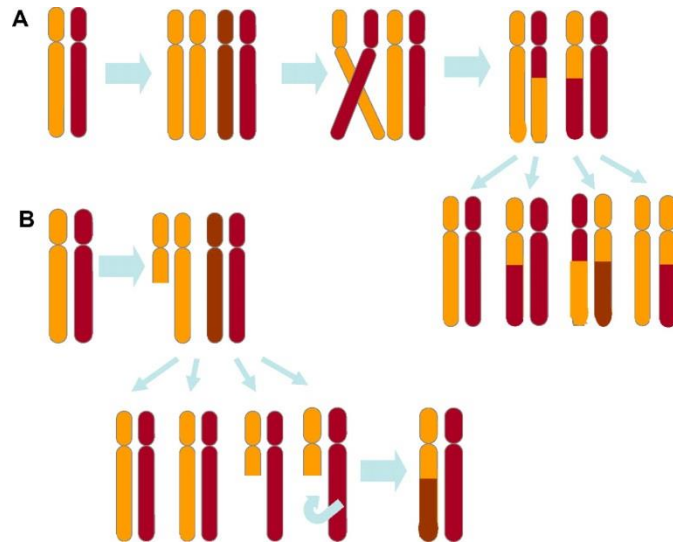
Figure 3.3 Mitotic mechanisms producing LOH regions. (Keefe *et al.*, 2010) **a)** mitotic recombination between homologous chromosomes, which yield daughter cells with different chromosome conformations. **b)** Deletion of one region can produce LOH by recombination with the homologous chromosome.

3.3.2   Somatic mutation identity verification *in silico*

To identify the putative chromosome breaking point, we searched for the first SNP position heterozygous in 'UFO-4' and homozygous in 'UFO-4Mut'. This was at position Pp6: 24,236,564 . The closest heterozygous SNP upstream was placed 5Kb apart. Following, we designed a primer pair to amplify the region including both polymorphisms (Pp06:24231526..24236686). After amplification and sequencing of some internal fragments we validated them. The DNA used was derived from fruit flesh (LII) to avoid chimeric DNA. This region contains part of a gene (*Prupe6.G244900*) that encodes for an uncharacterized protein. No transposable or repetitive elements were identified in this region.

The extended region with reduced heterozygosity identified *in silico* was also validated by amplifying and sequencing short DNA regions (using DNA from fruit flesh) distributed every 1Mb (approximately) along the 6.5 Mb affected region in the original and mutated samples. In total we designed primer pairs to target 7 regions with heterozygous/homozygous SNPs detected *in silico* (Figure 3.4). Sequencing results confirmed that all *in silico* SNPs were real, heterozygous in the original 'UFO-4' and homozygous in the mutant.
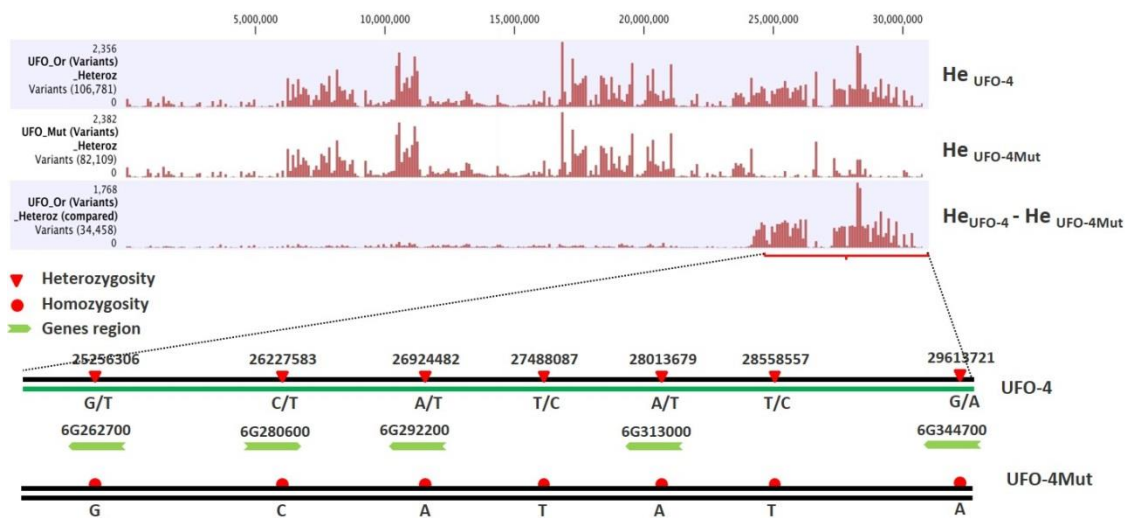
Figure 3.4 Overview of SNP polymorphisms in 'UFO-4' and 'UFO-4Mut' along the LOH region. The upper three images depict heterozygotes in 'UFO-4' and 'UFO-4Mut', and heterozygotes comparison in the two (i.e heterozygous loci in 'UFO-4' not shared with 'UFO-4Mut'). Bars height is proportional to the number of heterozygous polymorphisms. The bottom image is a schema graph showing SNP alleles in 'UFO-4' ad 'UFO-4Mut' along LOH region. Most of these SNPs loci are located within transcripts which are indicated in green boxes.

Figure 3.5 Sequence electropherogram showing the alleles of the SNP at position Pp06: 26,924,482 in heterozygosis in 'UFO-4' (upper line) and in homozygosis in 'UFO-4Mut' (bottom line). Black arrows indicate SNP position. Additional electropherogram pictures are shown in Supplementary Figure 3.S1.

3.3.3 A preliminary FISH method for checking peach somatic mutation

To validate the hypothesis of DSB followed by break repair with the homologous chromosome leading to a LOH region, we opted for the visual detection of the affected region through fluorescence *in situ* hybridization (FISH) technique. For this we designed three probes along the affected arm of the chromosome. These probes were designed in regions with large variants heterozygous in 'UFO-4' and homozygous in 'UFO-4Mut'.

Therefore, the probes should hybridize only in the chromosome carrying the round *s* allele (homozygous also in 'UFO-4Mut') (see cartoon in Figure III.6). One of the probes was designed inside the 10 Kb deletion affecting the flat candidate gene (*Prupe6.G281100*) and the other two 11.1 Kb upstream and 17.7 Kb downstream. Probes were obtained by PCR amplification of genomic DNA fluorescence labeled in the green channel signal.
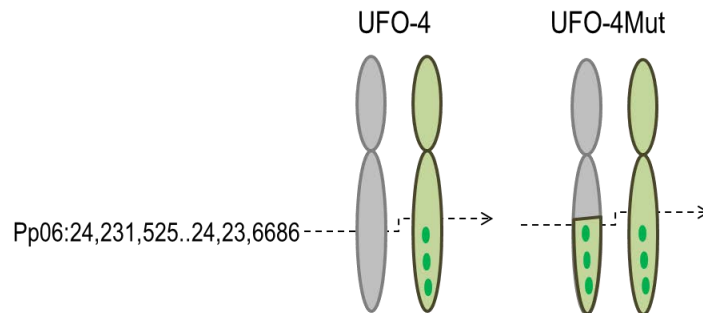
Figure 3.6 Cartoon showing probe hybridization expected in chromosomes from 'UFO-4' and 'UFO-4Mut' cells.

Chromosomes in metaphase were obtained from root rips (where cell division is more active) of 'UFO-4' and 'UFO-4Mut' plants previously introduced in *in vitro* culture. We corroborated that roots from 'UFO-4Mut' carried the genome rearrangement by PCR using Flatin1/Kinase5 primers (see chapter 1). After a 4 hour of pre-treatment in darkness of root tissue in a mixture of 8-hydroxyquinoline and colchicine we obtained high quality of chromosomes in metaphase. A following fixation step stabilized chromosomes for further hybridization and remove undesired cell products from the preparation. The next significant procedure was tissues digestion with an enzymatic treatment. In this step, a cell-wall lytic medium softened and digested the cell wall which can interfere with the FISH hybridization procedure inhibiting target sequence detection. This enzymatic treatment yielded cell suspension with little cytoplasmic debris remains. Drops of cell suspension were visualized in a high-resolution microscope.

To validate the quality of the chromosome preparation we visualized samples of the cell suspensions in a high-resolution microscope. To do so we prepared slides by dropping, fixing and staining small amounts of the cell suspension onto prechilled microscope slides. With DAPI staining, we analyzed the dynamics of chromosome spreading by visualizing chromosome structure with and epiflourescent microscope. Results revealed low number of the cells in metaphase although with good chromosome spreading. The low number of cells in the required conditions conditioned the success of the technique. Image of cell preparations are shown in Figure 3.7.

Despite the low number of cells in metaphase we attempted chromosome hybridization with the probes designed. Therefore, denatured fluorescence labeled probes were added to the microscope slides, allowing its binding with their homologous

DNA sequence on the chromosomes under conditions for renaturation. Hybridized slides were placed under an epifluorescent microscopy with the appropriate filters in the green channel. Unfortunately, we did not observed signal, indicating a miss-hybridization of the probes.
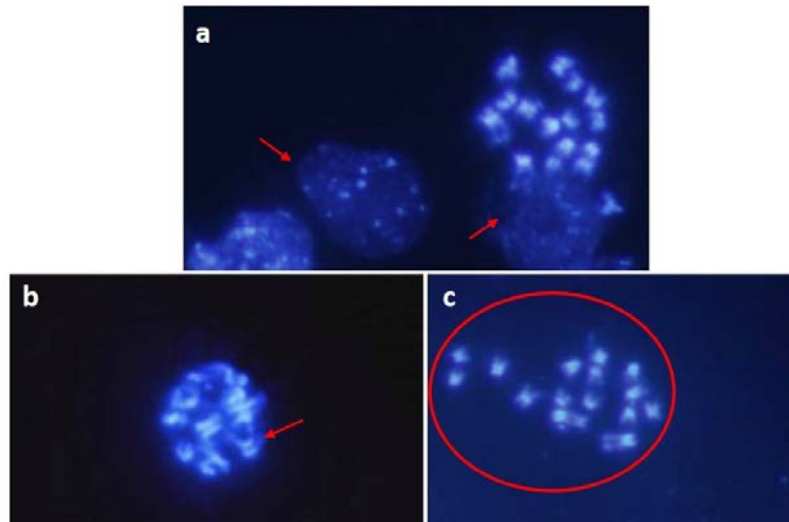
Figure 3.7 Cell suspension prepared with chromosomes in metaphase. **a)** Images show some cellular debris contaminating the preparation (red arrows). **b)** Chromosomes with un-digested cell way and cytoplasm pointed by red arrow. **c)** Chromosome preparation indicated by a red circle shows good chromosome spreading for FISH but with a relatively low yield. All chromosomes were stained with DAPI and the images were obtained in 100x objective under the epifluorescent microscope Axiophot DP70 with the DAPI filter (395-440 nm).

Chapter 4: PPV virus-based construction for gene function validation

4.1 Background

Motivated by (i) the availability of infectious *Plum pox virus* (PPV) constructions (López-Moya et al. 2000), (ii) the report of its use in infecting peach cultivars (the rootstock 'GF305' in particular) (Lansac et al. 2005), (iii) the use of virus constructions as vectors to express genes in peach (ref) and (iv) the need of developing a tool for gene validation, during this thesis we attempt to develop an infectious construction of a PPV clone carrying the peach fruit shape candidate gene *Prupe.6G281100*. The objective was to develop a tool to express a given gene, using *Prupe.6G281100* gene as example.

4.2 Materials and methods

4.2.1 Viral vector and bacterial strains

The construction was cloned within an initial *Plum pox virus* (PPV) viral vector pSNPPV5'BD-GFP which is 26,091 bp long and consists of an antibiotic (kanamycin) resistance gene, a fluorescent reporter gene and some other genes encoding functional virus proteins. In detail, it contains: functional, full-length cDNA copy of PPV RNA with the enhanced Cauliflower mosaic virus (CaMV) 35S promoter and nopaline synthase (NOS) terminator; tens of nucleotides downstream a NeoR/KanR antibiotic resistance insertion; total *Aequorea forskalea* GFP 717 nt (238 amino acid) was inserted between NIb and CP cistrons. This constructed virus can express fluorescent signal as a free protein, which could be used for detecting PPV movement in plants. The rest of pSNPPV5'BD-GFP has approximately 16 Kb length containing a Kanamycin resistance gene 18,339- 19133 bp (Supplementary Sequence 4.S1). This plasmid was obtained from Juan Antonio Garcia's lab (CNB, CSIC, Madrid) and stored at -20°C for subsequent use. This PPV-based plasmid cDNA was used as backbone to generate construction with foreign genes. In the present study, the *Escherichia coli DH5α* made in our lab was selected as competent cell for cloning purposes. The electroporation method with a high efficiency of up to 10^9 - 10^{10} transformants/μg DNA was used for plasmids transformation (Ryu & Hartin, 1990). And *Agrobacterium tumefaciens* GV3101 and EHA 105 strains were used for plasmids inoculation into target plants. A plasmid designed to express the P1b silencing suppressor was used as negative control (i.e., P1b is a duplicated P1 coding sequence with RNA silencing activity in the *Potyviridae* family) (Valli *et al.*, 2006). Since the presence of this potent RNA silencing suppressor allows robust expression of co-agroinfiltrated constructs, we anticipated its use to enhance the chances of infection with recombinants PPV-derived variants.

4.2.2 Plant materials

Model research species *Nicotiana benthamiana* and susceptible PPV host plant *Prunus persica* (peach) cv. 'GF305' were used to be inoculated under greenhouse conditions. To obtain plant materials for virus inoculation, enough amount seeds of *Nicotiana benthamiana* were sowed by sprinkling on the surface of substrate potting mixture. Keep moist until germination occurring within 21 days. After germination, plants were cultivated at 25°C with a 16-h-day/8-h-night cycle. Peach cultivar 'GF305' is widely used as commercial rootstock. This cultivar has been used in various experiments to evaluate *Prunus* resistance to PPV based on its extreme susceptibility (Martínez-Gomez *et al.*, 2000). In this study, one year 'GF305' seedlings were maintained in IRTA greenhouse at Torre Marimon (Caldes de Montbui, Barcelona) and transferred to the CRAG greenhouse for virus inoculation.

4.2.3 GFP-tagged PPV/foreign gene construction

4.2.3.1 Gibson assembly to construct recombinants

To make the GFP-tagged PPV/foreign gene construct, Gibson Assembly seamless cloning method was used for conjugating pieces of fragments (Gibson, 2009). Those segments to be assembled were obtained in separate PCRs using the same pSNPPV5'BD-GFP as template for virus-derived regions. Primers used for the amplifications were complementary to one another by covering nucleotides at their 5' ends that are complementary (overlapping sequences) to the 3' portion of the other primer to the corresponding flanking fragments. As a consequence, the PCR amplicons shared homologous sequences at the ends to be assembled. Unique-site restriction endonuclease enzymes *XbaI* and *BamHI* were selected to obtain the linearized backbone vector. The inserted gene *Prupe.6G281100* (name simplified here as PA11m) with a size of 2,253 bp was amplified from peach genomic from young leaves (more details about the foreign gene can be found in the first chapter of this document and in (López-Girona *et al.*, 2017)). Three PCR fragments plus one vector in each new recombinant were mixed and assembled with a Gibson Assembly master mix (New England Biolabs, Massachusetts, US), yielding the constructed recombinants (Figure 4.1). Detailed information of primers designed for three types of construction are reported in Table 4.1.
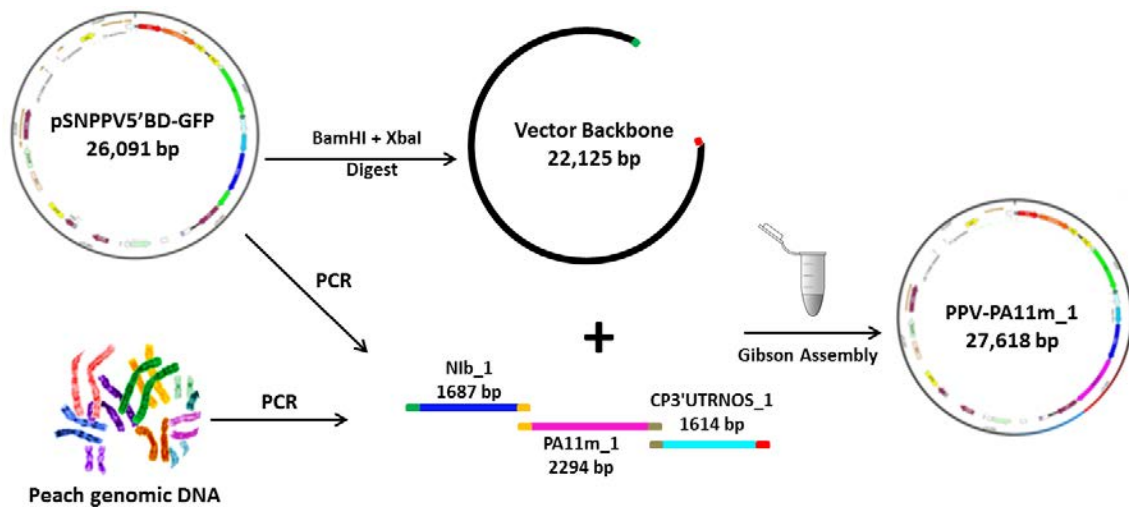
Figure 4.1 Schematic representation of Gibson Assembly to construct recombinant PPV_PA11m_1. Fragment backbone was obtained from double digestion on initial pSNPPV5'BD-GFP; fragments NIb-1, CP3'UTRNOS_1 were obtained by tailed primers on pSNPPV5'BD-GFP; fragment PA11m_1 was obtained by tailed primers on peach genomic DNA. The final ligation reaction was carried out in a micro-tube following assembly protocol.

4.2.3.2 GFP-tagged PPV/foreign gene recombinant characterization

The final plasmids constructed were named as PPV-PA11m_1, PPV-GFP-PA11m_2, PPV-PA11m–GFP_3 that were characterized by agarose gel running after PCR, enzymes digestion and Sanger sequencing. To confirm plasmid integrity, overlapping primers NIb_c123_fwd, CP_c123_rev, and primers inside candidate gene Prupe.6G281100 were used to amplify specific bands. Molecular weight and restriction endonuclease-generated fragments pattern for the plasmids were calculated from the agarose gel electrophoresis by EcoRI and PstI endonucleases cleavage, using HindIII-EcoRI digested λ-DNA as size standard. In order to check if there were any sequence mutations or errors during the processes of amplification, recombination and assembly of fragments, all the recombinants were partially sequenced at CRAG for their characterization.

Table 4.1 Primers and fragments used for three types of recombinants. Fragments with various sizes in different recombinant were named correspondingly. Lowercase letters in primer sequences are from one conjoint fragment with overlapping parts, while uppercase letters cover the other adjacent fragment.

Recombinant	Fragment	Length	Primer	seq 5'- 3'
PPV-PA11m_1	NIb_1	1,687 bp	NIb_c123_fwd	tacaacccggatgaagtttgctggGGATCCCTACAACTCAAGAGG
			NIb_c13_rev	gcaaatgtttcatGGCCTGATGTACTACGAC
	PA11m_1	2,294 bp	PA11m_c13_fwd	agtacatcaggccATGAAACATTTGCTCCAATATTTC
			PA11m_c12_rev	ctctttcgtcagcttgatgaaccactacgttACGTCTCCACCTTC
	CP3'UTRNOS_1	1,614 bp	CP_c12_fwd	agtggttcatcaaGCTGACGAAAGAGAAGACG
			CP_c123_rev	ttgcatgcctgcaggtcgactctagAGATCTAGTAACATAGATGACACCGCGCGA
	Vector	22,125 bp		
PPV-GFP-PA11m_2	NIb-GFP_2	2,436 bp	NIb_c123_fwd	
			NIbGFP_c2_rev	gcaaatgtttcatAGCTTGGTGCACAACAAC
	PA11m_2	2,294 bp	PA11m_c2_fwd	tgtgcaccaagctATGAAACATTTGCTCCAATATTTC
			PA11m_c12_rev	
	CP3'UTRNOS_1	1,614 bp	CP_c12_fwd	
			CP_c123_rev	
	Vector	2,2125 bp		
PPV-PA11m-GFP _3	NIb_1	1,687 bp	NIb_c123_fwd	
			NIb_c13_rev	
	PA11m_3	2,294 bp	PA11m_c13_fwd	
			PA11m_c3_rev	tgctcaccatggcttgatgaaccactacgttACGTCTCCACCTTC
	GFP-CP3'UTRNOS_3	2,352 bp	GFPCP_c3_fwd	agtggttcatcaagccATGGTGAGCAAGGGC
			CP_c123_rev	
	Vector	22,125 bp		

4.2.4 Agro-inoculation on plants and GFP visualization

4.2.4.1 Plants agro-infiltration

The initial plasmid pSNPPV5'BD-GFP isolate was propagated by transformation into *Escherichia coli DH5α* following heat shock protocol: mix well 50 µL of competent *DH5α* cells and 2 µL of plasmid in a 1.5 mL Eppendorf; incubate the mixture on ice for 30 min, transfer to water-bath at 42°C for 45 s; move onto ice and keep for 2 min; add 950 µL LB medium; incubate at 37°C with strong agitation 250 rpm for 1.5 h; centrifuge at 5000rpm for 5 min; remove most of the supernatant, leaving 100 µL that are used to

re-suspend the pellet; spread the re-suspended cells on LB plates containing Kanamycin to select the resistant clones, and leave the plate at 37°C overnight. The following day, colonies were picked and selected by PCR with adequate primers to identify the positive ones that were used for plasmid extraction according to plasmid mini-prep protocol (GeneJET™, Thermo Scientific). Prepared plasmids were then transformed into the *A. tumefaciens* strains *GV3101* and *EHA105*. Both strains were cultivated for further inoculation (named agro-inoculation) on the model species *Nicotiana benthamiana* and on the targeted plant *Prunus persica* (peach) cv. 'GF305'. Leaves of several individual plants of *N. benthamiana* and 'GF305' where agro-infiltrated with the virus construct and with a mock solution on years 2016 and 2017. For this approximately 250 μL of the infiltration mixture (from a culture reaching an optical density at 600 nm of 0.5 for each strain) were applied with a syringe to the underside of two leaves of *N. benthamiana* or "GF305" peach plants. Infiltration reagent was composed of 10 mM 2-(N-morpholino) ethanesulfonic acid (MES, pH5.5), 10 mM $MgCl_2$, 150 μM acetosyringone (López-Moya, 2000).

4.2.4.2 GFP detection and imaging

After 15 days of post-inoculation (dpi), GFP fluorescence was monitored under a stereoscope epifluorescence microscope (Olympus SZX16) using GFP and GFPA filter (excitation filters at 460/95 BA510IF and 460/95 BA510-550, respectively) and photographed with an Olympus DP71 digital camera (12.5 megapixels) at Cell D software. When photographing with Cell D, the image condition was set as maximum opening of the diagram by using ISO 1600 sensitivity in 30% spot size. After an exposure time of 6.75 seconds, the high live quality image was snapshot and saved for analysis.

4.3 Results

4.3.1 Infectivity test of pSNPPV5'BD-GFP

In this work we used a plasmid pSNPPV5'BD-GFP based on *Plum pox virus* (*PPV*) which was constructed for vector and GFP expression (Supplementary Figure 4.S1). This plasmid was designed with CaMV 35S promoter regulating transcription of viral PPV cDNA which was followed by NOS terminator cistron. The construct provides Kanamycin resistance for selection of plants after stable transformation, but for inoculation with the virus only transient expression (without selection) was required: the infection process was expected to proceed systemically from the agro-inoculated leaves to invade the whole plant taking advantage of the cell-to-cell and long-distance movement functions of PPV in the susceptible hosts. The GFP sequence was inserted between NIb and CP cistrons by using a duplicated polyprotein cleavage site, producing a final functional vector pSNPPV5'BD-GFP able to monitor the movement of the virus. To construct GFP-tagged PPV/gene recombinants (containing both the GFP and the gene-

of-interest, with further cleavage sites), the plasmid backbone and the required fragments were cloned and multiplied in *E. coli DH5α* using a Gibson assembly strategy. Each *E. coli* culture carrying plasmid pSNPPV5'BD-GFP and derived variants was selected on Kanamycin containing LB-agar plates. After purification of plasmids and mobilization into A. tumefaciens, leaves of *N. benthamiana* and 'GF305' plants were agro-inoculated with the virus constructions, as well as with a mock solution as negative control. Weak virus-related symptoms, consisting on vein yellowing and light green rings were observed often in both species approximately 5-7 days of post-inoculation (dpi). These weak symptoms were occasionally accompanied by dark areas (likely caused by anthocyanin accumulation) that appeared progressively in the adjacent leaves. After 15 dpi, both infectious symptoms and GFP-associated fluorescent signals could be visualized in upper young leaves of both *N. benthamiana* and *P. persica* plants. Thus, these results confirmed that the initial GFP-tagged PPV was infectious and capable of systemic movement in susceptible plants.
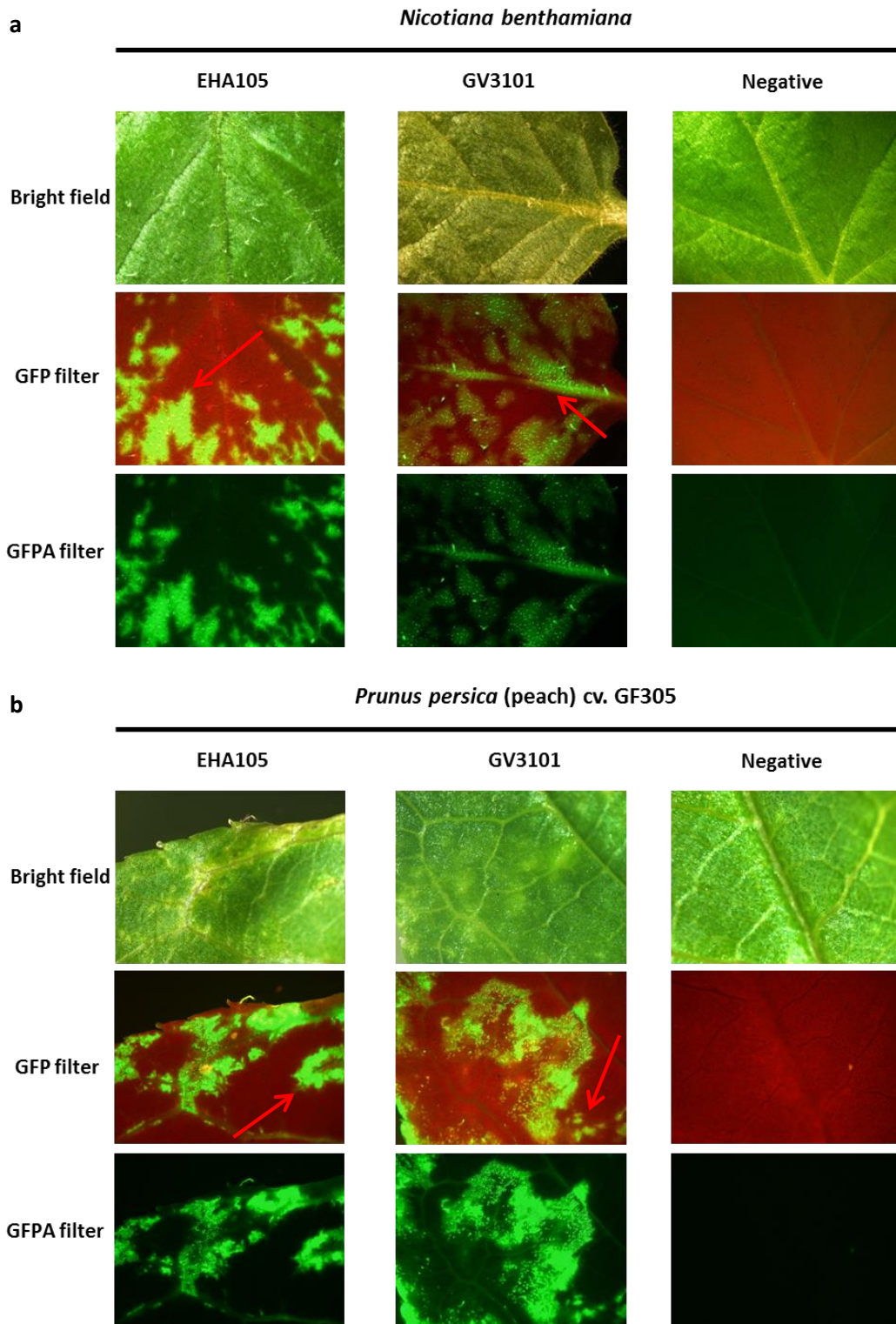
Figure 4.2 GFP fluorescence and symptoms details of leaves infected with pSNPPV-5´BD-GFP in N. benthamiana **a)** and P. persica **b)** plants. Both plants were agro-inoculated with Agrobacterium tumefaciens GV3101 and EHA 105 strains, plasmid PPV with P1b silencing suppressor was selected as negative for blank control. GFP and GFPA filters were used for observing fluorescent GFP signals which are indicated with red arrows.

4.3.2 PPV vector constructions carrying *Prupe.6G281100*

The plasmid vector used in this analysis (pSNPPV-5´BD-GFP) has a size of 26,091 bp. The introduction of an additional gene with 2,253 bp in such a large backbone could compromise its stability and therefore its functionality. Considering this possibility, we designed three different constructions for PPV-based plasmids considering the relative position of the candidate gene (*Prupe.6G281100*, simplified as PA11m) and the GFP gene: PPV-PA11m_1 (c1) in which the GFP region was replaced by PA11m, PPV-GFP-PA11m_2 (c2) with the GFP upstream of the PA11m, and PPV- PA11m-GFP_3 (c3) with the GFP downstream of the PA11m.

For the c1 construction, we designed a procedure to delete the 717 bp green fluorescent reporter (GFP) while simultaneously inserting the foreign gene. This could be implemented by using *Prupe.6G281100* genomic DNA cloned from a round peach (i.e. the functional allele) and amplifying each linearized fragment from the original plasmid pSNPPV5'BD-GFP with overlapping primers. The backbone vector containing 22,125 bp was linearized by double digestion of pSNPPV5'BD-GFP using *XbaI and BamHI* enzymes. One overlapping primer pair forward (NIb_c123_fwd) incorporated a *BamHI* site, which in combination with a reverse primer (NIb_c13_rev) produced a fragment (NIb_1) with 1,687 bp incorporating a NIa cleavage site for NIb cistron, resulting in a motif sequence NVVVHQ/A. PA11m_1 genomic DNA was cloned from a round peach sample using overlapping primers combination PA11m_c13_fwd + PA11m_c12_rev. The terminal stop codon TGA was expurgated (to assure the proper expression of the polyprotein) and followed again by a synthetic protease site NVVVHQ/A, producing a 2,294 bp fragment. Overlapping primers CP_c12_fwd and CP_c123_rev amplified a 1,614 bp fragment CP3'UTRNOS_1 from pSNPPV5'BD-GFP. All these four fragments could be assembled into a final viral recombinant, named PPV-PA11m_1 with a length of 27,618 bp (Figure 4.3).

The construction c2 (PPV-GFP-PA11m_2, Figure 4.3) was designed to integrate PA11m downstream the GFP, and upstream the CP region. In total, four modified fragments were required to assemble this recombinant with a total length of 28,362 bp, where the backbone vector and fragment CP3'UTRNOS_1 were the same as described above. A primer pair NIb_c123_fwd + NIbGFP_c2_rev was used to get fragment NIb-GFP_2 with 2,436 bp. While the candidate gene, 2,294 bp PA11m_2 from peach genomic DNA was amplified by overlapping primers PA11m_c2_fwd + PA11m_c12_rev (obtain the candidate gene with discriminative tags for various constructions). A new synthetic cleavage site NVVVHQ/A was included while removing its own terminal stop codon TGA to assure the continuity of the viral polyprotein.

The third construction c3 (PPV- PA11m-GFP_3, Figure 4.3) consisted of the insertion of PA11m downstream of the NIb region, and upstream of the GFP. The final

product with 28,362 bp contained same backbone and NIb_1 cistrons, and novels PA11m_3 and GFP-CP-3'UTRNOS_3. Fragment PA11m_3 was a 2,294 bp PCR amplicon obtained by primers PA11m_c13_fwd + pPA11m_c3_rev from peach DNA (obtain the candidate gene with discriminative tags for various constructions), whose terminal stop codon TGA was expurgated and substituted by a synthetic cleavage site NVVVHQ/A. Primes GFPCP_c3_fwd and CP_c123_rev was used to amplify fragment GFP-CP-3'UTRNOS_3, generating a 2,352 bp fragment.
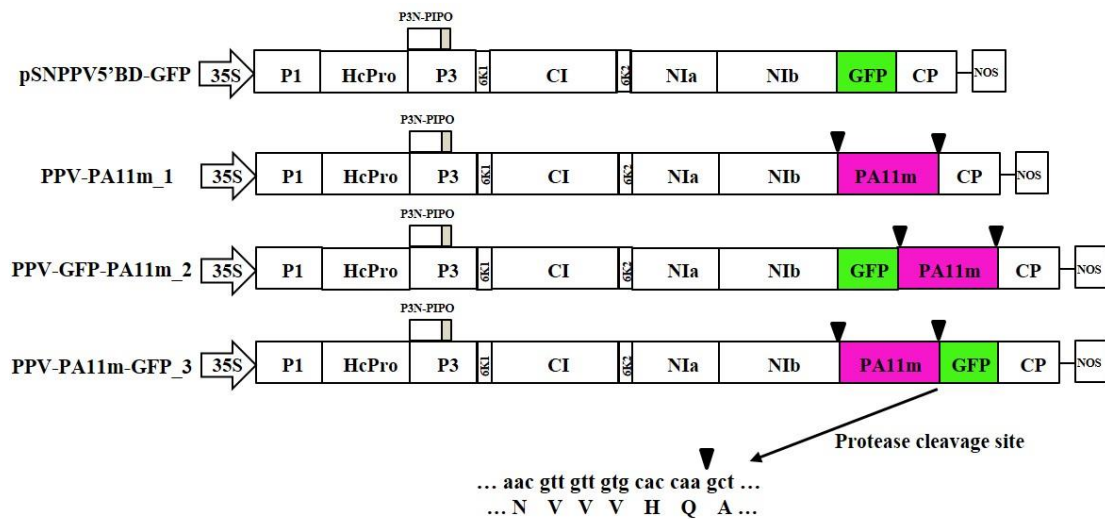
Figure 4.3 Schematic representation of PPV full-length cDNA clones and the new recombinants. Only the synthetic protease cleavage sites located flanking the foreign inserted gene PA11m (Prupe.6G281100) are depicted (as inverted black triangles). PPV functional cistrons present in the constructs are shown with their names in the different boxes, the green one represents the fluorescent protein GFP, and the pink box is the newly insert candidate gene.

4.3.3 Characterization of PPV variant recombinants containing *Prupe.6G281100*

After all the three plasmids had been assembled under *in vitro* condition, they were characterized by performing restrictions with endonuclease cleavage sites by two selected enzymes. Figures showed the agarose gel electrophoresis patterns of PPV-PA11m_1, PPV-GFP-PA11m_2, PPV-PA11m-GFP_3 that had been treated with restriction endonucleases. *The HindIII-EcoRI* digested λ-DNA was used as a standard for visualizing fragment sizes. An image showing the restriction endonuclease-generated fragments was analyzed to verify the recombination construction strategy (Figure 4.4).

Both *EcoRI* and *PstI* endonucleases could cleave the three recombinants generating diagnostic fragments for the presence of PA11m. For *PstI*, the pattern of restriction digestion of construct PPV-PA11m_1 will be distributed into 5 segments: a 11,043 bp fragments containing the Kanamycin resistance gene and CaMV 35S promoter, a 9,435 bp diagnostic fragment containing most of the PPV cistrons and candidate gene

Prupe.6G281100, a 4,923 bp fragment, a 1,947 bp fragment containing NOS terminator, as well as a 270 bp fragment. The *EcoRI* endonuclease produced 13 segments: The Kanamycin resistance gene and NOS terminator were included in fragment 12,583 bp, others were 3,747 bp, 1,918 bp, 1,647 bp, 1,595 bp, 1,513 bp, 1,125 bp, 1,005 bp, 804 bp, 690 bp, 508 bp, 423 bp, 60 bp. The inserted *Prupe.6G281100* was recognized by the presence of diagnostic fragments. In PPV-GFP-PA11m_2, the *PstI* recognition sequence generated similar pattern as previous PPV-PA11m_1. The only difference was that the second large fragment contained 10,179 bp with GFP inside. Likewise, in *EcoRI* digestion 13 fragments were obtained having the third large fragment 2,339 bp. This was a new segment conjugated with GFP inside, replacing the fifth 1,959 bp fragment in the previous pattern. The endonuclease-generated fragments pattern of *PstI* recognition sequence sites in PPV -PA11m-GFP_3 was same as that of PPV-GFP-PA11m_2. But there were two different fragments expected for *EcoRI* digestion: a new 1,548 bp instead of the previous 804 bp. As shown in Figure 4.4, our results confirmed the presence of the expected restriction fragments.
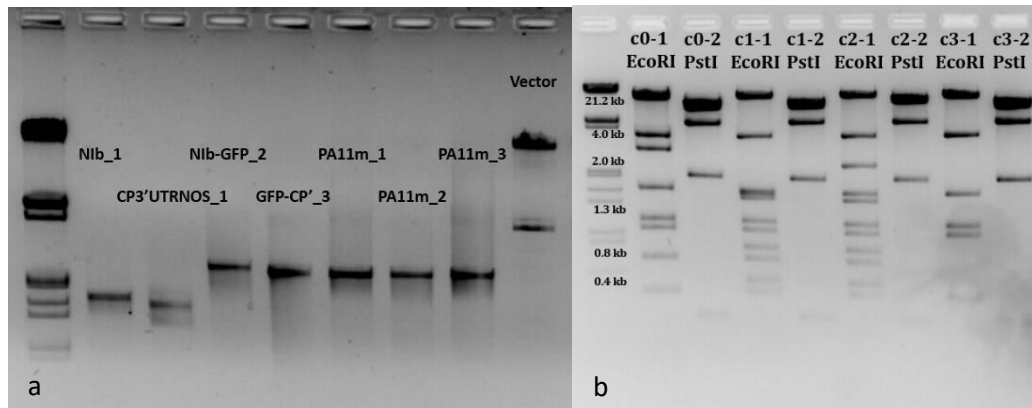
Figure 4.4 Agarose gel to characterize recombinants. **a)** Fragments prepared for Gibson Assembly were visualized on agarose gel, and all the expected bands were cut and purified afterwards. **b)** Positive colonies were characterized by endonuclease EcoRI and PstI. Plasmid c0 is the initial pSNPPV5'BD-GFP, while c1 equals to PPV-PA11m_1, c2 is PPV-GFP-PA11m_2, c3 is PPV-PA11m-GFP_3. In each plasmid, two colonies were picked and verified. The HindIII-EcoRI digested λ-DNA was used as a standard for visualizing fragment size.

4.3.4 Plant inoculation assay

The two constructions including the GFP gene were agro-infiltrated in both *N. benthamiana* and *P. persica* plants. Contrasting with the observations when infecting with the pSNPPV5'BD-GFP plasmid, we did not observe virus symptoms in any of the agro-inoculated plants. To evaluate if the infections could be asymptomatic, we collected leaves at 15 dpi and at 21 dpi and observed for the presence GFP fluorescence (Figure 4.5).

Only weak signals of fluorescence were present in the infiltrated leaves of the model plant *N. benthamiana*, showing a sort of disperse distribution around the injected area, and near some minor veins (Figure 4.5 a). However, GFP signal was hardly observed in young leaves, appearing in a sporadic distribution which were inconclusive for the presence of the virus. These results suggested that the GFP-tagged constructs with the PA11m insertion were either non-infective or severely impaired to generate detectable infections. Indeed, the weak signals could derive from an inefficient direct translation of the construct without virus replication.

Even worse, in the susceptible genotype of *P. persica* 'GF305', we failed to observe any fluorescent signals in neither the infiltrated nor in the young leaves (Figure 4.5 b). These observations indicated that the GFP-tagged PPV recombinants were probably unable to move from cell to cell and systemically in the peach seedlings. The difference of behavior with the assays in *N. benthamiana* plants might reflect further troubles to infect a woody host, compared with the highly virus-permissive herbaceous model plant.

The construction without GFP (c1) was not used for plant inoculation, due to the difficulties created by the absence of any visually detectable marker that could help us to detect virus replication (requiring for instance highly sensitive specific RT-PCR methods) and will be tested out of this thesis.
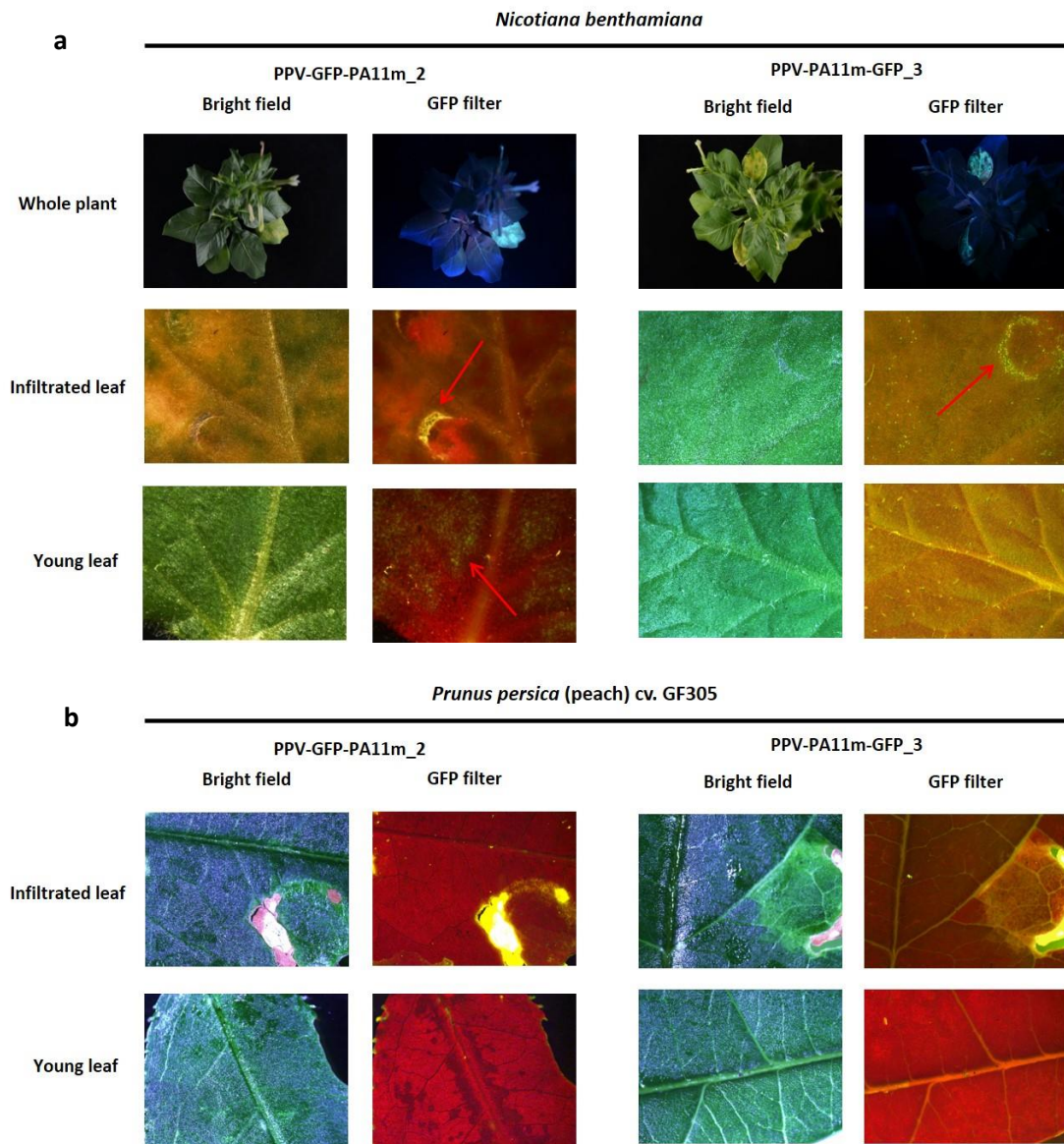
Figure 4.5 GFP fluorescence and symptoms details of leaves infiltrated with GFP-tagged PPV recombinants PPV-GFP-PA11m_2 and PPV -PA11m-GFP_3 in *N. benthamiana* **a)** and *P. persica* **b)** plants. All the plants were agro-inoculated with Agrobacterium tumefaciens GV3101 strain. In the whole plant, infiltrated leaf, and young leaf level, GFP filters were used for observing inconclusive fluorescent GFP signals which are indicated with red arrows.

General discussion

Peach (*Prunus persica*) is a model species for genetic and functional genomics studies on fruit trees principally because of its small and simple genome and comparatively short juvenile period (Horn *et al.*, 2005; Arús *et al.*, 2012). Peach fruit, which derives from the mature ovary, shows abundant phenotypic variability in agronomic traits, such as those involved in organoleptic quality like aroma volatile compounds (Eduardo *et al.*, 2010; Bianchi *et al.*, 2017), sugar, acidity content and firmness (Byrne *et al.*, 1991; Cantín *et al.*, 2009; Cirilli *et al.*, 2016) as well as in those traits related with fruit appearance like skin pubescence (peach/nectarine), fruit flesh and skin color (yellow, white and red) and fruit shape (including flat and round as well as intermediate phenotypes). Intense research activity studying the genetics of those peach fruit traits is currently providing valuable knowledge on the genetic mechanisms controlling them and promote the development in peach marker assisted selection (MAS) breeding. The ultimate objective of such studies is to identify the genes, and its allelic variants, responsible for the observed variation on trait. In this aspect, peach genetics nourishes from research conducted in model organisms like *Arabidopsis thaliana*, tomato, maize or rice. For example, results published by Hauser *et al.* reporting the function of the *GLABROUS1* (GL1) R2R3-MyB transcription factor gene in trichome formation in *Arabidopsis* (Hauser *et al.*, 2001) supported the postulation by Vendramin *et al.* of the R2R3-MyB gene *PpeMYB25* as the causal gen for skin trichomes in peach (Vendramin *et al.*, 2014). Similarly, the mention of the peach NAC transcription factor encoded by the gene *ppa007577* as responsible for fruit ripening is strongly supported by its homology with the *NOR* gene in tomato (Giovannoni *et al.*, 1995; Osorio *et al.*, 2011).

This doctoral thesis focuses on the peach fruit shape trait, aiming at understanding its genetic determination and developing tools for candidate gene validation. In the first chapter of this thesis we identify a candidate gene for flat shape (*Prupe6.G281100*). As in the above examples, the homology of the proposed gene (a Leucine Repeat Receptor Like Kinase; LRR-RLK) with *Arabidopsis* genes involved in meristem development, strongly supported its involvement in the trait determination. LRR-RLKs are highly expanded in plants as cause of genome duplications (Fischer *et al.*, 2016). Although peach genome has not been duplicated (Verde *et al.*, 2013), some gene clusters are observed. Among them we want to mention a cluster of LRR-RLKs on the fruit shape locus. The study of four plant genomes (*Arabidopsis*, grape, poplar and rice) showed that LRR-RLK genes expanded in clusters show high ratios variability (probably induced by positive selection (Tang *et al.*, 2010). Gene duplication and retrotransposition occurred during the expansion process may generate changes in the DNA of the gene disabling its function and therefore (Xiao *et al.*, 2016). These non-functional genes derived from functional genes are called pseudogenes, and despite that they do not produce functional RNA or protein may be involved in gene regulation (Sasidharan & Gerstein, 2008). A nonfunctional copy of *Prupe6.G281100* is postulated

here as the causal of the flat shape of peach; this gene accumulates large levels of variability. In addition, a close observation of the genes in its proximity (also LRR-RLK genes) reveals some other genetic variants. This, together with the discovery of three peach samples where their genotype at *Prupe6.G281100* does not explain their fruit shape, we started studied in depth the variability in this cluster. This analysis is presented in chapter 2. The third and four chapters aimed at developing tools for candidate gene validation with the practical example of the candidate gene postulated in chapter one. Peach is recalcitrant for genetic transformation. Here we intended to develop useful approaches or techniques at genome and chromosome level, including the whole-genome sequence analysis of a sport mutant, fluorescent in *situ* hybridization (FISH) and virus-based vector construction.

Identification of a candidate gene for flat shape in peach

(This section has been adapted from (López-Girona *et al.*, 2017))

In the first chapter we explored the genetic variability in a region associated with the flat shape in peach and identified a candidate gene for this trait. Considering the high level of variability observed genome-wide between round and flat peaches (Aranzana *et al.*, 2003) the large extension of LD in peach (Aranzana *et al.*, 2010; Li *et al.*, 2013) and the codominant mode of action of the flat and round alleles, which must be in heterozygosis in flat varieties, we expected to find a substantial level of heterozygosis in the region flanking the SSR marker associated with the trait (UDP98-412). Surprisingly none of the fragments sequenced showed polymorphisms. Thereafter we searched for the closest region to UDP98-412 with annotated SNPs in the databases. This region was 337.5 Kb upstream and contained one SNP every 521 bp, close to the density of 1 SNP every 598 bp found by (Aranzana *et al.*, 2012) after sequencing genes in peach varieties, but much higher than the density of 1 SNP every 1076bp observed by (Cao *et al.*, 2014) in Chinese edible varieties.

After sequencing nine amplicons of the variable region in a panel of varieties we identified SNPs highly associated with flat, round and aborting phenotypes in two amplicons of the gene *Prupe.6G281100*. By amplification, cloning and sequencing part of the gene, 11 SNPs and two INDELs co-segregating with the trait were identified (Figure 1.1), which allowed us to design an allelic specific marker diagnostic for this trait. This marker was validated in 177 varieties from different origins, including nineteen where UDP98-412 alleles escaped association with the trait (Picañol *et al.*, 2013). In all cases, the genotype obtained was in agreement with fruit shape phenotype, confirming that this region is closer to the S/s locus. In consequence, we provide here a simple marker (FlatIn_F/IndelS_F; PC4) able to amplify two fragments differing in 5 bp, that improves the performance of UDP98-412 and is more efficient for MAS. Additionally, to this primer combination, we found several SNPs that can be

used for the same purpose. Long-range PCR reactions detected a ~10 Kb deletion of part of the gene (693 bp from the ATG starting codon), affecting the flat-associated allele. Alignment and coverage analysis of NGS reads of five flat and five round varieties allowed visualization of the alignment of the large gap in heterozygosis (Figure 1.3) This deletion was validated in the panel of varieties, with all flat varieties sharing the same haplotype and suggesting a unique origin of the flat trait in the panel evaluated. Some of these varieties (18) have been analyzed in (Micheletti *et al.*, 2015) with an SNP array of close to 9,000 SNPs. They were distributed along all transects of the variability observed including, the major Oriental and Occidental clusters, indicating that the varieties analyzed here covered a broad range of variability. Unlike peaches and nectarines that are separated in different clusters in Occidental materials (Aranzana *et al.*, 2010), no specific clusters including only flat peaches occur, which is consistent with the fact that, due to the heterozygous nature of flat peaches, breeding is usually by selecting in round x flat progenies, and that the flat allele, originating from a single source, may have been introgressed in a diverse array of materials.

RT-PCR (Figure 1.4) and posterior band sequencing revealed the absence of transcription of the flat associated allele, indicating a loss of function of *Prupe.6G281100*. Thereafter *Prupe.6G281100* gene, from the orthologous group ORTHO03D000261, annotated as *BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1* (*BAK1*) arises here as a candidate for the flat shape of peach fruit. This gene is a leucine-rich repeat receptor-like kinase (LRR-RLK), the proteins constituting ligand-receptor systems that control cell fate specification, and mediate correct cell divisions and cell-to-cell communication, allowing correct generation of tissues and organs through growth and development of both animals and plants (Cock *et al.*, 2002). Plant RLKs can be classified into six classes based on the structural feature of the extracellular domain. The largest class of plant RLKs is the LRR-RLKs class (700 in *Arabidopsis* and 1,400 in rice) (Matsushima & Miyashita, 2012), proteins that contain leucine-rich repeats, which are tandem repeats of approximately 24 amino acids with conserved leucine involved in protein-protein interactions. Most LRR-RLKs are involved in embryonic pattern formation, which suggests a putative role of this protein in the coordination of cell proliferation during embryogenesis and during morphogenesis of embryonic cells at meristems, shaping the plant (De Smet *et al.*, 2009). Two LRR-RLKs, *CLAVATA1* and *ERECTA*, show functional implications in the maintenance, size and shape of meristems (Mandel *et al.*, 2014).

The protein showed best homology with *BAK1*, *FLORAL ORGAN NUMBER1* (*FON1*) and *THICK TASSEL DWARF1* (*TD1*) genes. *BAK1* is involved in brassinosteroid (BR) signal transduction, forming heterodimers with *BRASSINOSTEORID INSENSITIVE* (*BRII*) modulating growth and development, including cell expansion and reproductive development in species such as *Arabidopsis* and rice (He *et al.*, 2007; Zhang *et al.*, 2014). The *FON1* gene encodes a receptor-like kinase protein (orthologous to

Arabidopsis CLAVATA1) that regulates the size of the floral meristem, causing enlargement in *Oryza sativa* (Suzaki *et al.*, 2004). Similarly, *TD1* encodes a maize orthologous to *CLAVATA1* in *Arabidopsis*, modulating meristem size during inflorescence and flower development and involved in the regulation of meristem structural organization (Bommert *et al.*, 2005). We can therefore hypothesize that the LRR-kinase protein encoded by *Prupe.6G281100* is involved in a cell signaling pathway, during flower development, that ensures a final round shape of the ovary, and consequently of the fruit. While the loss of function of this gene in homozygosis produces unviable fruit, in heterozygosis the allele produces flat fruit. This behavior resembles the mechanism of a haploinsufficient locus. Loss-of-function alleles at haploinsufficient loci are typically dominant because the level of gene function in a heterozygote is below the threshold for producing a wild-type phenotype, and homozygotes typically exhibit more severe phenotypes, including early lethality. The most common explanation is that these loci are involved in cellular processes sensitive to dosage effects and changes in protein concentration (Birchler & Veitia, 2010).

Our candidate gene differs from that suggested by (Cao *et al.*, 2016), identified through a GWAS approach. These authors found associated SNPs in the fifth intron of a *CONSTITUTIVELY ACTIVATED CELL DEATH GENES* (*CAD1*) homologous gene, which negatively controls the salicylic acid (SA) mediated pathway of programmed cell death in plant immunity. This gene is 650 Kb downstream from our candidate gene. Flat varieties contained the polymorphism in either homozygosis (A/A) or heterozygosis (A/T), while round varieties were always homozygous T/T, indicating that these genotypes do not fully correspond with the inheritance of the trait (A/A genotypes should not produce viable fruit). To our knowledge, Cao *et at.* study is the first to report a putative role of *CAD1* genes in organ shape and development and its high association with the trait could be due to the large LD extension in peach. However, we cannot discard possible involvement of both genes in the trait.

Genetic variability exploration in LRR-RLK genes cluster of the *S* locus

In the candidate gene homology analysis (chapter 1) we found several homologous genes in the peach genome, which, as suggested for *Arabidopsis* (He *et al.*, 2007) might be functionally redundant. Given that most of its homologues clustered in the same genome region of the *S* locus, the large LD in peach and the unusually high variability detected in the haplotype, we need to explore other possible functional polymorphisms in the region acting alone or in combination.

In addition, the routinely use of the markers designed to genotype the 10 Kb polymorphisms of *Prupe.6G281100* in marker assisted selection (MAS), identified three samples for which the genotype did not predict the phenotype: M24, M247 (both from a breeding program, the second first-generation offspring of the first) and

'Cerrito', a Brazilian peach canning variety. In particular cultivars M14 and 'Cerrito' show an oblate shape (between flat and round) with of the allele carrying the 10 Kb deletion in heterozygosis; while M247 present flat shape with a haplotype of 10 Kb deletion in homozygosis.

Whole genome resequencing of the three cultivars (chapter 2) has revealed two large polymorphic regions in *Prupe.6G281400* (15.3 Kb downstream *Prupe.6G28110)*, and in *Prupe.6G281500* (21.3 Kb downstream *Prupe.6G28110).* While the first polymorphism was observed only in the outlier varieties, the second was linked to the 10 Kb in *Prupe.6G28110.* Large fraction of the polymorphisms consisted on SNPs and small indels in phase with the flat associated allele in *Prupe6.G261500* respect to the reference genome. SNP validation by Sanger sequencing revealed that most of the *in silico* polymorphisms in *Prupe.6G281400* were caused by the misalignment of a highly polymorphic allele of *Prupe.6G281500*. Indeed, sequence misalignment due to high levels of allelic variation is a common event during both *de novo* assembly or reads alignment to a reference genome with the use of NGS technologies (Narzisi & Mishra, 2011; Robinson *et al.*, 2011). We consider interesting that the polymorphic allele of *Prupe.6G281500* aligned with *Prupe6.G281400* gene, with which it shared 87% of nucleotides. This fact reinforces the hypothesis of possible redundant function of the LRR-RLK genes of this locus. Therefore, although we cannot affirm that *Prupe6.G281500* is involved in fruit shape determination, the full linkage of its polymorphic allele with the 10 Kb deletion in *Prupe6.G281100* determines that this gene still needs to be considered into fruit shape determination. In addition, we are here categorizing fruit shape as flat and round but intermediate shapes are also observed in peach fruit, including the oblate one. Therefore, we cannot discard that the interaction between these genes determines different fruit shape types.

In this sense, M14 shows a clear oblate phenotype while 'Cerrito' shape tends to be also oblate when grows in cold areas ('Cerrito' is a low chilling cultivar, i.e. requires low number of chilling hours for flowering). In warm areas 'Cerrito' has a round shape. Contrary to *Prupe.6G281500*, *Prupe.6G281400* showed a large deletion shared only by these three outlier cultivars. It consists on a 6.1 Kb deletion affecting the upstream region starting 192 bp before the 5'UTR of the primary transcript and ending at the end of the 5'UTR, just 144 bp before the start codon of the gene. In addition, we observed high polymorphism flanking this deletion. The polymorphisms upstream (indels and SNPs) were in phase with the 10 Kb deletion in *Prupe.6G281100* while the ones downstream occurred only in the allele carrying also the 6.1 Kb. Therefore, we may conclude that the 6.1 Kb deletion in *Prupe.6G281400* occurred in the flat associated allele, it is to say, this deletion is younger than the 10 Kb deletion and both are in phase. The 6.1 Kb deletion stands a good chance of affecting important regulatory elements together with *Prupe.*6G281400 promoter, giving a much shorter intergenic region.

To predict *Prupe.6G281400* transcription start site (TSS), we used an online tool *Softberry TSSPlant* that was developed to search RNA polymerase II promoters (transcription start sites, TSSs) in plant DNA sequences, of a wide spectrum of plant genomes (Shahmuradov *et al.*, 2017). Consequently, a putative positive chain TSS position was detected a few base pairs following the 6.1 Kb deletion start, in which 30 bp upstream with a predicted TATA box (Figure D.1) specifically binding transcription II D (TF$_{II}$D) (Lee & Young, 2000). This preliminary prediction agreed with the assumption that transcription regulatory elements (REs) are located in the highly conserved region, which accordingly consists with the fact that the transcription factors (TFs) are conserved group of genes, whereas kinases and transporters remain variable to adapt rapidly to the changing environmental conditions (Spivakov *et al.*, 2012; Tatarinova *et al.*, 2016). From Figure 2.9 the transcription start site (TSS) and parts of the promoter region bound to REs, essential for being recognized by RNA-polymerase II (Orphanides *et al.*, 1996), are missing due to the 6.1 Kb deletion. In addition, it cannot be neglected that a part of the 6.1 Kb non-coding region contains class I transposons which jump off at some points during breeding programs or in wild populations, leaving a large deletion and SNPs surrounding the breaking point. Likewise, intronic transposable elements (TEs) insertion affecting transcript stability was suggested to be recognized in yellow flesh peach allele alteration (Falchi *et al.*, 2013).
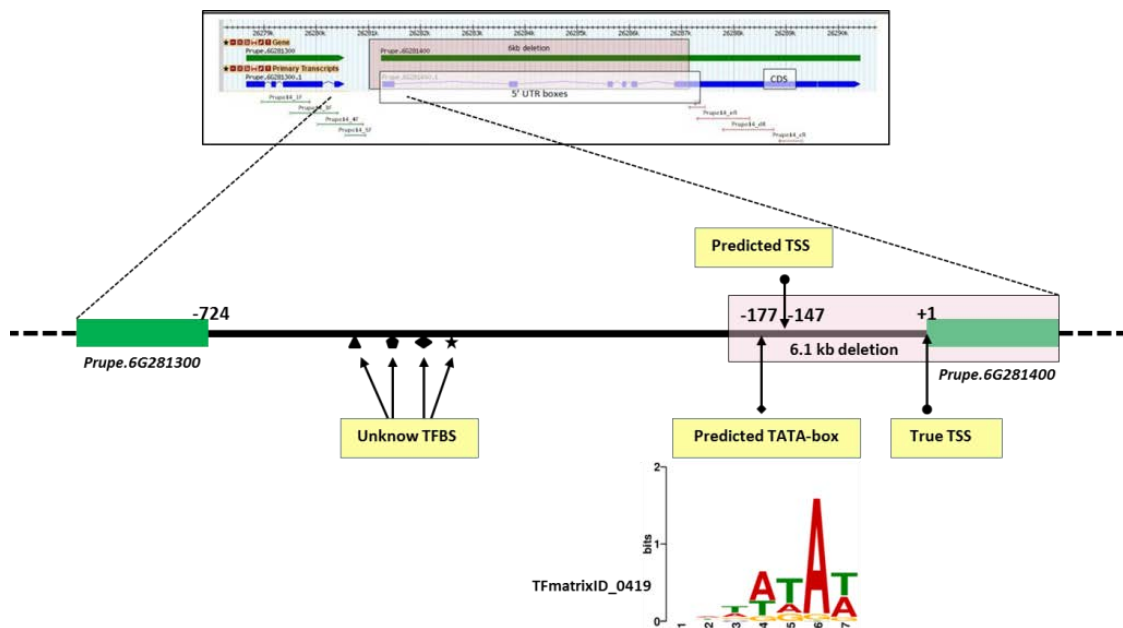
Figure D.1 Schematic representation of transcription start sites (TSS) prediction for *Prupe.6G281400*. Green boxes are parts of transcripts in peach genome, while the big deletion upstream in pink box. Predicted transcription factor binding site (TFBS) TATA-box, transcription start site (TSS) and other unknown TFBS are indicated with yellow boxes and black arrows. Numbers showing relative position based on Prupe.6G281400 are symbolled with "+" (from start to right) and '-' (from start to left). Graphical sequence log exhibits a transcription II D (TF$_{II}$D) sequence TFmatrixID_0419 binding correspondent to TATA-box (Hahn *et al.*, 1989; Chow *et al.*, 2015).

By Gene Ontology (GO) study (Gene Ontology Consortium, 2004), both of the two genes are annotated as binding protein (GO:0005515), including gene *Prupe.6G281500* as well. The InterPro classification system (Hunter *et al.*, 2009) describes them as identifiers IPR032675, IPR001611, IPR013210 within homologous superfamily leucine-rich repeat like kinases (LRR-RLKs, leucine-rich repeat-containing N-terminal, plant-type). Moreover, Gene function prediction analysis shows that the three LRR-RLKs function have best homology with *BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1* (*BAK1*), *FLORAL ORGAN NUMBER1* (*FON1*) and *THICK TASSEL DWARF1* (*TD1*). These genes have been identified to be involved in mediating floral meristems and organs growth during inflorescence and flower development, including and structural organization, shape alteration and size regulation in various species (Suzaki *et al.*, 2004; Bommert *et al.*, 2005; Moon *et al.*, 2006; Postel *et al.*, 2010).

Our RNA analysis did not detect transcription of both genes when carrying their respective deletions. Therefore, the lack of function in *Prupe.6G281100* (due to the 10 Kb deletion) and *Prupe.6G281400* (due to the 6.1 Kb deletion) together with *Prupe.6G281500* polymorphic allele may be implicated by interacting together, in combination or alone in peach fruit shape determination. However, we do not consider the transcription analysis of *Prupe.6G281400* conclusive since we did not used combination of primer to reverse-transcribe all the coding regions of the gene. In case of transcription, sequence data predicts that the *Prupe.6G281400* allele carrying the 6.1 Kb deletion would encode a polypeptide, losing 614 amino acids accompanying a lack of most of LRR motifs and domains, the transmembrane domain and the uncertain C-t regions.

Here we hypothesize that the studied LRR-RLKs encoded by *Prupe.6G281100, Prupe.6G281400 and Prupe.6G281500* should account for a round peach ovary shape by regulating cell signaling pathway during flower development. Our data suggest a three-allele model hypothesis for fruit shape determination, where each allele is determined by the combination of polymorphisms at these three LRR-RLK and which present different levels of dominance. The allele with higher dominance (*$*) is the one accumulating the 10 Kb deletion in *Prupe.6G281100*, the 6.1 kb deletion in *Prupe.6G281400* and the polymorphic allele in *Prupe.6G281500*. Cultivars carrying this allele in either homozygosis or heterozygosis show round or oblate shape, respectively. The second allele in terms of dominance (*S*) is the flat allele reported in bibliography. This allele carries the deletion in *Prupe.6G281100* and the polymorphic allele in *Prupe.6G281500* but has intact the *Prupe.6G281400* gene. This allele is dominant over the third allele (*s*) which carries all three LRR-RLK functional (Table D.1). In addition, aborting and oblate shapes in function of the allele combination suggest a dosage effect of them in fruit development. This reinforces the haploinsufficiency model postulated in previous paragraphs since haploinsufficient gene requires both alleles to be functional in order to express the wild type (round shape in this case). When an event

like a sequence deletion results in a loss-of-function allele at haploinsufficient loci, the gene expression level in a heterozygote is down-regulated and is insufficient to produce the wild-type phenotype, showing intermediate phenotypes; homozygotes might present more severe phenotypes, including altered or even fatal disease (Seidman & Seidman, 2002; Kondrashov & Koonin, 2004). The combination of this deletion with a second in a homologous linked gene shows a reversion of the phenotype to the wild type. Therefore, gene *Prupe.6G281400* might be involved in an inverse cell signaling pathway and act as interaction with gene *Prupe.6G281100.* The function lack of *Prupe.6G281400* counteracts the function lack of *Prupe.6G281100*, leading to fruit shape alteration in cultivar M14, M247 and 'Cerrito'. This hypothesis needs further validation by studying gene-gene interaction as well as the level of expression of the genes in the signaling pathway.

Table D.1 Hypothesis for fruit shape phenotype determined by the possible allele combinations. **$** allele combines deletion in *Prupe.6281100* and in *Prupe.6281400* as well as the polymorphic allele in *Prupe.6281500*. Allele **S** combines the deletion in *Prupe.6281100* and the polymorphic allele in *Prupe.6281500*. **s** has all genes functional as in the reference genome. Grey intensity represents dominance intensity.

	$	S	s
$	Round	oblate	Round
S	Oblate	Aborting	Flat
s	Round	Flat	Round

Study of a peach somatic mutant with round shape derived from a flat fruit variety

Gene function is usually validated by genetic transformation or by the screening of mutants. Currently, the transformation and transgenic regeneration in peach has been developed by using embryo-derived callus and embryo axis section from seeds (Smigocki & Hammerschlag, 1991; Pérez-Clemente *et al.*, 2005; Padilla *et al.*, 2006), but this approach still remains an obstacle for peach genes study because of time-consuming procedure and low regeneration rate of transformed plantlets (Scorza *et al.*, 1990).

Alternatively, the study of somatic mutants in woody plants, and in particular in peach, has been successfully used to investigate causal genes (Brandi *et al.*, 2011; Falchi *et al.*, 2013). These mutations often occur in only one histogenic layer, so are

chimeric and most are not sexually transmitted (Foster & Aranzana, 2018). In peach, the histogenic layer LI gives rise to epidermal tissues, LII to subepidermal tissues, and the male and female sporogenous tissues, and LIII to the remainder of the shoots. In fruit, LI produces the skin, LII the flesh and LIII the stone.

Here (Chapter 1) we investigated the *Prupe.6G281100* gene in a chimeric natural mutation occurring in the meristematic LII (producing the fruit flesh tissue), which reverted from the flat to the round phenotype (Figure 1.5). The analysis of flesh DNA with allele specific primers to genotype the *Prupe6.G281100* gene revealed a new structural mutation affecting the flat allele, while the skin DNA shows the intact flat and round-associated alleles. One hypothesis for the gain of function compatible with the haploinsufficiency mechanism is the recombination of the mutant flat allele with others of the LRR-Kinase genes present in the candidate gene region. As demonstrated by (Albert *et al.*, 2010), chimeric kinase receptors made in the lab can produce new functional receptors. In fact, sequence divergence, genetic recombination, duplication events and selective forces have been proven to be the main forces for the continuous RLK gene expansion (Rodgers-Melnick *et al.*, 2012). Alternatively, chromosome replacement of the 'flat' region by the homologous 'round' region is also a plausible hypothesis.

To clone the allele in the mutant sample (Chapter 3), we resequenced genomic DNA from leaves of the original flat variety ('UFO-4') and of its round mutant ('UFO-4Mut'). To overcome the chimerism of the leaf tissue we sequenced both samples at high depth (higher than 90x). Indeed, bioinformatic analysis of DNA sequences obtained from chimeric tissue is a challenge because the sequence reads from the altered chromosome are usually under-represented. Recently Marroni *et al.* developed a bioinformatic pipeline to identify genome-wide structural changes in chimeric mutations by identifying regions with reduced levels of heterozygosity (ROH) (Marroni, Fabio *et al.*, 2017). Thanks to the high depth of the sequence reads and that the mutation occurred in the LII (the highest cell lineage represented in the leaf tissue) we could identify relevant levels of variation between 'UFO-4' and its clone in a large region affecting the distal arm of chromosome 6. In particular variation consisted in a considerable reduction of heterozygous polymorphism (i.e. polymorphisms in heterozygosis in the original were homozygous in the clone). We assumed and validated by PCR that chromosomes from LII cell lineage were completely homozygous in this region and therefore the reduced number of loci still in heterozygosis occurred only in the intact chromosomes of LI cell lineage. This region with loss of heterozygosity (LOH) span along a region of close to 6.5 Mb, starting at about 24.2 M position of the chromosome to its end. Polymorphisms analysis allowed determining the starting point of the chromosome rearrangement in a relatively narrow region (5 Kb). The large LOH region contains the *S* locus. The first implication of this finding is that, as the change affects a long region of the chromosome which including multiple genes, 'UFO-4Mut'

cannot be used to validate *Prupe.6G281100* gene as candidate for fruit shape. However, this mutant is still highly interesting to study mechanisms of somatic variability in peach.

The analysis of the sequence depth and polymorphisms along the LOH regions suggest that it may has been arise from the repair of a double strand brake (DSB) affecting the chromosome carrying the flat allele (*S*) with the homologous chromosome carrying the functional round allele (s). This fact well explains the reversion of the phenotype from flat (i.e. one nonfunctional copy of the gene) to round (with the two gene copies fully functional). These kinds of events have been frequently reported in cancer research (Ryland *et al.*, 2015) and with less frequency in plants. As wine grape breeding is based on clone selection, observed sports are often selected and clonally maintained for testing. In the research ambit, this provides an important genetic resource to elucidate mechanisms driving to somatic variability at the chromosome level. In grape, transposable elements (TE) have been reported as major cause of somatic variability (Carrier, Grégory *et al.*, 2012). A careful bioinformatic analysis of the 5 Kb region delimiting DSB point in 'UFO-4Mut' discarded TEs as the cause of the rearrangement. A part of TE, rearrangement events produced by random DNA breaks have also been reported (Torregrosa *et al.*, 2011; Carbonell-Bejerano *et al.*, 2017). In particular, chromosome replacement and/or deletion of one of the two copies of the gene regulating anthocyanin pathway in 'colored' Pinot noir produced Pinot blanc 'white' clone. As a consequence, the affected region showed a loss of heterozygosis (LOH) in the white mutant (Walker *et al.*, 2006; Yakushiji *et al.*, 2006; Pelsy *et al.*, 2015). Pelsy *et al.* demonstrated, the 'white' haplotypes of Pinot blanc come from the replacement of different sections of the 'colored' haplotype by its 'white' homolog (both present in Pinot noir) (Pelsy *et al.*, 2015).

Although at a different level than in grapevine, somatic mutations in peach have also been historically relevant for plant breeding. As most of commercial traits are controlled by single or major genes (Dirlewanger *et al.*, 2006; Arús *et al.*, 2012), somatic mutation easily will involve phenotypic (and therefore visible) changes. Currently, large number of commercial peach varieties are sports (Okie, 1998) and few of them have been studied to validate gene function. To our knowledge research, works to unravel genetic mechanisms driving somatic changes in peach have not been reported. Therefore, the study of changes occurred in 'UFO-4Mut' provides relevant and novel knowledge. However, to determine how often genome rearrangements are involved in peach somatic variability, additional peach sport mutants should be studied.

Despite the highly conclusive data obtained from the bioinformatic analysis of the sequences complemented by wet-lab genotypic data, we intended to develop a strategy to visually detect the chromosome replacement by means of fluorescence *in situ* hybridization (FISH) technique. In other plant species, chromosome screening using somatic cell preparations has been proved to be efficient for exploring chromosomal

abnormalities, such as chromosomal termini transgenes from chromosomal truncations in maize (Yu *et al.*, 2006), structural rearrangements of hexaploid tritordeum detection for mitotic chromosome spreads in *Triticum* (Cabo *et al.*, 2014), intro- and interspecies polymorphisms identification and comparison of individual chromosomes in *Hordeum* and *Triticum aestivum* (Rey *et al.*, 2018). However, most of these FISH detection is related to hybridization between single-gene or loci and specific probes with limited size, which is different from mutant chromosome visualization in peach reported in this dissertation.

In peach, the mutant clone contains a non-canonical 'round' haplotype with 6 M of LOH where the candidate genes *Prupe.6G281100* and *Prupe.6G281400* within fruit shape *S* locus are involved. This might stem from somatic recombination or mitotic gene conversion based on recombinational repair of chromosomal double strand breaks (DSBs) event. During this repair procedure, one of the haplotype which needs to be repaired acts as a receptor to duplicate the non-reciprocal transfer of information from the other homologous donor sequence, producing a new mutant clone with non-canonical donor haplotype (Moynahan & Jasin, 1997; Symington, 2002). This molecular mechanism agrees with our case in which a non-canonical 'round' haplotype receives a canonical one, producing mutant peach spot with round fruit phenotype.

During the setting of the protocol we have identified several critical steps in chromosome preparation that need to be tuned for successful hybridization of chromosomes and probes in future. The first step is to select the right tissue with high rate of cells in division. Initially we started with bud meristems and leaf tissues (experiments not reported in this thesis), which resulted in a very number of cells in metaphase. Therefore, we decided to use root tips, usually used in the FISH analysis reported in the bibliography. 'UFO-4' and 'UFO-4Mut' trees were grafted in a rootstock, therefore we needed generate roots from them. For this we introduced these two plants in *in vitro* culture from tender branches with internodes, and root them. Roots derive from LII cell linages and therefore are convenient to identify mutations occurred in LII but not those in LI or LIII. The second key step for obtaining enough number of adequate chromosomes in metaphase division is the treatment to arrest cells in metaphase stage by inhibiting the formation of functional spindles during mitosis (Taylor, 1965). Here we used a combination of 8-hydroxyquinoline and colchicine to improve the number of cells in metaphase; this combination provides better results in number of metaphases and clearness than the use of only each of them separately (Tlaskal, 1979). Tissue fixation (which stabilizes the tissue and removes undesired cell materials from the preparation) and its enzymatic digestion (which softens and digests the cell wall to avoid its interference with the FISH hybridization), are also critical to provide cell suspensions cleaned of cell debris adequate for probe hybridization. In our protocol these two steps remain unsolved since our preparations contained an excess of cell debris that may have prevented from probe hybridization (Figure 3.7). In deed tuned up for this process will

include to investigate the appropriate quality of fixative (freshly/long-playing, cold/room temperature) as well as the different ratio of methanol and acetic acid required (Sumner *et al.*, 1973; Islam & Levan, 1987), to choose the right enzymes and to set the optimum digestion time.

Our attempt here represents a first step in the implementation of FISH technique for the study of somatic rearrangements in peach chromosomes. We are confident that further tune up trials will allow obtaining clean and good spread metaphase chromosomes for cytological study of mitotic cells in peach and other *Prunus* species.

Virus-based vector construction used for gene study in peach fruit shape

Gene function in plant organisms can be identified or validated in various ways among which genetic transformation is widely used to find out how the target genes regulate the corresponding phenotypes (Shah *et al.*, 1987; Walden & Schell, 1990). Usually transformation relies in the overexpression or silencing of genes. However, there are some species that are recalcitrant for genetic transformation. This is the case of peach, being plant regeneration the main challenge.

Virus-induced gene silencing (VIGS) and virus induced expression appear as an alternative for modifying gene product levels in peach. This method has been already reported in peach. For example, Jia *et al.* used a *Tobacco rattle virus* (TRV) vector to silence *PpCHLH* gene in peach. By including a cDNA fragment of that gene, chlorophyll biosynthesis was reduced in peach leaves (Jia *et al.*, 2015).

Aiming at developing a similar virus-based tool for expression or silencing genes in peach, here (Chapter 4) we initiated the development of viral constructs able to infect peach. For these we constructed recombinants between an infectious full-length clone plasmid of *Plum pox virus* (PPV) containing GFP e, and a peach gene (in this case the candidate for flat shape *Prupe.6G281100*). The virus chosen was PPV because of the high infectivity and mobility in peach of certain strains. The peach variety selected was 'GF305', described to be highly susceptible to PPV. 'GF305' is a rootstock frequently used in commercial orchards. 'GF305' fruit are round; therefore, we do not expect fruit shape changes in infected plants (overexpressing the gene). However, thanks to the mobility of the virus through the plant, flat and aborting cultivars grafted over this rootstock should experience such shape changes.

The plasmid vector that we used (pSNPPV5'BD-GFP) was developed by the Juan Antonio Garcia group to visualize, thanks to the fluorescence generated by the expression of the GFP gene, virus localization and movement in the plant. We took advantage of the availability of this construction to start with the development of a virus tool to express genes in peach. Initially we validated the infectivity of the basic construction in peach. First yellowing symptoms were observed few days after infection.

Virus movement from infected leaves to new young leaves was validated by fluorescence observation at 15 dpi. Similar symptomatic time was observed by Díaz-Vivancos *et al.* in pea plants using a PPV-GFP vector (Díaz-Vivancos *et al.*, 2008).

The functional plasmid used contains a t-DNA region with a Kanamycin resistance gene followed by a full-length copy of the genome of PPV under the transcriptional control of a strong CaMV 35S promoter. This positive-sense ss-RNA genome encodes a 340-370 KDa polyprotein comprising P1, HCPro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP (from N- to C-terminus). The mature polyprotein is processed by three virus-encoded proteases: P1 and HCPro carry their own cis-cleavage sites by the respective P1 serine proteinase and HC-Pro cysteine proteinase at their C termini (Verchot *et al.*, 1991; López-Moya *et al.*, 2009), while NIa-Pro cysteine proteinase accounts for cis- and trans-cleavages of the remaining gene products located in the C-terminal two-thirds of the polyprotein (Carrington & Dougherty, 1987). In our pSNPPV5'BD-GFP, the reporter gene GFP was inserted between NIb and CP cistrons, incorporating a site that could be also recognized by NIa-pro to separate GFP from NIb and CP.

Initially we set a strategy to insert the whole gDNA of the flat shape candidate gene into the PPC-GFP vector down- and upstream GFP (constructions c2 and c3, respectively), between the NIb and CP regions. Since CP is one of the most variable regions among the potyviral gene products and its integrity could compromise infectivity (Shukla & Ward, 1988; Bousalem *et al.*, 1994), a correct cleavage of the inserted gene product was assured by incorporation of the required cleavage sites. Complete GFP and *Prupe.6G281100* genes could be successfully cloned and validated by digestion with *EcoRI* and *PstI* endonucleases (Figure 4.4). In addition, the constructions were validated by Sanger sequencing.

The two recombinant plasmids with GFP and PA11m were agro-infiltrated in *N. benthamiana* and peach plants to evaluate their infectivity and, in affirmative case, for further gene function validation. Few days after infection, *N. benthamiana* plants showed faint symptoms in the infiltrated leaves, however the virus presence was not confirmed in *N. benthamiana* young leaves nor in peach leaves, where fluorescence failed to be detectable. There are several possible reasons for this failure: it could be caused simply by the large size of the construction (10 Kb of the virus+ 2.2 Kb of the gene), which could exceed the limits tolerated by a potyvirus genome to be able to replicate efficiently. Therefore, to mitigate the increase of size we prepared a construction without GFP. The construction procedure as well as the validation method was as in the previous constructs, but in this case, it was not possible to detect the virus with any fluorescent marker. This construction has not been agro-infiltrated yet, and future work will be needed to test if it is infective or not. In fact, at this point we cannot rule out other hypothesis, like interference of PA11m with virus replication or movement. Again, experiments to verify replication in protoplasts would be needed to

address this, and probably new control constructs will be desirable, for instance with other foreign gene products.

In summary, here we have started the development of a virus-mediated tool for gene expression in peach, that will require further refinements to be useful. This tool may be highly relevant in the analysis of new gene functions in plants like peach with long regeneration time, and recalcitrant for genetic transformation.

Conclusions

1. Deep characterization of DNA and RNA analysis of the gene *Prupe.6G281100* reveal that this gene is not transcribed in flat peaches.

2. This lack of transcription is probably due to the large deletion (close to 10 Kb) affecting its promoter and coding region.

3. Lack of transcription of *Prupe.6G281100* is compatible with a haploinsufficient mechanism in fruit shape determination.

4. Track visualization of next generation sequence (NGS) reads has been proved to be useful to identify large structural variants as well as misaligned regions.

5. *Prupe.6G281100* clusters with other four highly homologous LRR-RLKs. Protein BLAST analysis found them to be close to genes which regulate stem cell population size, floral meristem and organ grown during inflorescence and organ development.

6. The flat associated allele in *Prupe.6G281100* is linked to additional variants in *Prupe.6G281500*, placed 23.1 Kb apart.

7. Some varieties escape the association between *Prupe.6G281100* and *Prupe.6G281500*, suggesting the involvement of different alleles in the peach fruit shape determination.

8. A 6 Kb deletion affecting *Prupe.6G281400* (15.3 Kb apart from *Prupe.6G1100*) in phase with the flat associated allele explain the phenotype of the few varieties escaping the association.

9. RNA analysis of the allele of *Prupe.6G281400* gene carrying the 6 Kb deletion suggest that it is non-functional.

10. We have developed a three allele model hypothesis that explains the phenotype. This hypothesis needs to be validated in a larger sample panel.

11. The LRR-RLKs clustered in the *S* locus carrying non-functional allele(s) accumulate large number of SNPs and small variants compared to the functional LRR-RLKs genes in the cluster.

12. A double strand brake (DSB) of the chromosome carrying the flat allele repaired with the homologous chromosome (which carries the round allele) explains the round phenotype observed in the sport mutant of a flat variety.

13. Such repair was translated in a region 6.5 Mb long with loss of heterozygosity (LOH).

14. The large region, which involved thousands of genes, did not allow us to validate *Prupe.6G281100* function.

15. LOH regions as a result of genome rearrangements have been described in woody crops like grapevine, but to our knowledge this is the first time that is described in peach.

16. Somatic chimerism may challenge somatic variability analysis.

17. Fluorescent in situ hybridization (FISH) is a promising tool to visualize chromosome rearrangements in peach somatic cells. Metaphase arresting and enzymatic treatments are critical steps that need to be tune up for further use of this technique.

18. A fully viable plasmid based on *Plum pox virus* (PPV) and fluorescent reporter GFP was infectious in Nicotiana benthamiana and peach, indicating its viability as virus vector to attempt future constructions.

19. Despite the infectivity of the vector we could not obtain infective construction when combining with *Pupre.6G281100* gene sequence. We conclude that large the vector size could inhibit its infectivity or mobility.

20. Shorter construction lacking, for example GFP, may overcome this difficulty.

Bibliography

Abbott, A.G., Rajapakse, S., Sosinski, B., Lu, Z., Sossey-Alaoui, K., Gannavarapu, M., Reighard, G., Ballard, R.E., Baird, W.V., Scorza, R., & Callahan, A. (1998). *Construction of saturated linkage maps of peach crosses segregating for characters controlling fruit quality, tree architecture and pest resistance*.

Adami, M., de Franceschi, P., Brandi, F., Liverani, A., Giovannini, D., Rosati, C., Dondini, L., & Tartarini, S. (2013). Identifying a carotenoid cleavage dioxygenase (*ccd4*) gene controlling yellow/white fruit flesh color of peach. *Plant molecular biology reporter, 31*(5), 1166-1175.

Agusti, J., Lichtenberger, R., Schwarz, M., Nehlin, L., & Greb, T. (2011). Characterization of transcriptome remodeling during cambium formation identifies *MOL1* and *RUL1* as opposing regulators of secondary growth. *PLoS Genet, 7*(2), e1001312.

Aker, J., & de Vries, S.C. (2008). Plasma membrane receptor complexes. *Plant physiology, 147*(4), 1560.

Albert, M., Jehle, A.K., Mueller, K., Eisele, C., Lipschis, M., & Felix, G. (2010). *Arabidopsis thaliana* pattern recognition receptors for bacterial elongation factor Tu and flagellin can be combined to form functional chimeric receptors. *J Biol Chem, 285*(25), 19035-19042.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., & Nielsen, P.H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol, 31*(6), 533.

Albrecht, C., Russinova, E., Hecht, V., Baaijens, E., & de Vries, S. (2005). The *Arabidopsis thaliana* SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. *Plant Cell, 17*(12), 3337.

Alexander, L., & Grierson, D. (2002). Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. *Journal of Experimental Botany, 53*(377), 2039-2055.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research, 25*(17), 3389-3402.

Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schäffer, A.A., & Yu, Y.K. (2005). Protein database searches using compositionally adjusted substitution matrices. *The FEBS Journal, 272*(20), 5101-5109.

Aranzana, M.J., Abbassi, E.-K., Howad, W., & Arús, P. (2010). Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genetics, 11*(1), 69.

Aranzana, M.J., Carbó, J., & Arús, P. (2003). Microsatellite variability in peach [*Prunus persica* (L.) Batsch]: cultivar identification, marker mutation, pedigree inferences and population structure. *Theoretical and Applied Genetics, 106*(8), 1341-1352.

Aranzana, M.J., Illa, E., Howad, W., & Arús, P. (2012). A first insight into peach [*Prunus persica* (L.) Batsch] SNP variability. *Tree Genetics & Genomes, 8*(6), 1359-1369.

Aranzana, M.J., Pineda, A., Cosson, P., Dirlewanger, E., Ascasibar, J., Cipriani, G., Ryder, C.D., Testolin, R., Abbott, A., King, G.J., Iezzoni, A.F., & Arús, P. (2003). A set of simple-

sequence repeat (SSR) markers covering the *Prunus* genome. *Theoretical and Applied Genetics, 106*(5), 819-825.

Arús, P., Verde, I., Sosinski, B., Zhebentyayeva, T., & Abbott, A.G. (2012). The peach genome. *Tree Genetics & Genomes, 8*(3), 531-547.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics, 25*, 25.

Atanasoff, D., Dodoff, D.N., Kovacevski, I.C., Martimofp, S.I., Trifonova, M.V., & Christoff, A. (1932). Parasitic fungi new to Bulgaria. Third contribution. *Yearbook Univ. of Sofia. Fuc. of Agric., 10*, pp.341-366.

Bakir, Y., Eldem, V., Zararsiz, G., & Unver, T. (2016). Global transcriptome analysis reveals differences in gene expression patterns between nonhyperhydric and hyperhydric peach leaves. *The Plant Genome, 9*.

Bañuelos, M.A., Garciadeblas, B., Cubero, B., & Rodríguez-Navarro, A. (2002). Inventory and functional characterization of the HAK potassium transporters of rice. *Plant physiology, 130*(2), 784.

Batley, J., & Edwards, D. (2009). Mining for SNPs and SSRs using SNPServer, dbSNP and SSR taxonomy tree. In *Bioinformatics for DNA sequence analysis* (pp. 303-321): Springer.

Beauchemin, C., Bougie, V., & Laliberté, J.-F. (2005). Simultaneous production of two foreign proteins from a potyvirus-based vector. *Virus research, 112*(1), 1-8.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K.D., & Sayers, E.W. (2018). GenBank. *Nucleic Acids Res, 46*(D1), D41-D47.

Bianchi, T., Weesepoel, Y., Koot, A., Iglesias, I., Eduardo, I., Gratacós-Cubarsí, M., Guerrero, L., Hortós, M., & van Ruth, S. (2017). Investigation of the aroma of commercial peach (*Prunus persica* L. Batsch) types by proton transfer reaction–mass spectrometry (PTR-MS) and sensory analysis. *Food Research International, 99*, 133-146.

Birchler, J.A., & Veitia, R.A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol, 186*(1), 54-62.

Bishop, R. (2010). Applications of fluorescence *in situ* hybridization (FISH) in detecting genetic aberrations of medical significance. *Bioscience Horizons: The International Journal of Student Research, 3*(1), 85-95.

Blenda, A.V., Verde, I., Georgi, L.L., Reighard, G.L., Forrest, S.D., Muñoz-Torres, M., Baird, W.V., & Abbott, A.G. (2007). Construction of a genetic linkage map and identification of molecular markers in peach rootstocks for response to peach tree short life syndrome. *Tree Genetics & Genomes, 3*(4), 341-350.

Bliss, F.A., Arulsekar, S., Foolad, M.R., Becerra, V., Gillen, A.M., Warburton, M.L., Dandekar, A.M., Kocsisne, G.M., & Mydin, K.K. (2002). An expanded genetic linkage map of *Prunus* based on an interspecific cross between almond and peach. *Genome, 45*(3), 520-529.

Boch, J. (2011). TALEs of genome targeting. *Nat Biotechnol, 29*, 135.

Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120.

Bommert, P., Lunde, C., Nardmann, J., Vollbrecht, E., Running, M., Jackson, D., Hake, S., & Werr, W. (2005). *thick tassel dwarf1* encodes a putative maize ortholog of the *Arabidopsis CLAVATA1* leucine-rich repeat receptor-like kinase. *Development, 132*(6), 1235.

Boudehri, K., Bendahmane, A., Cardinet, G., Troadec, C., Moing, A., & Dirlewanger, E. (2009). Phenotypic and fine genetic characterization of the *D* locus controlling fruit acidity in peach. *BMC Plant Biol, 9*(1), 59.

Bouhadida, M., Moreno, M.Á., Gonzalo, M.J., Alonso, J.M., & Gogorcena, Y. (2010). Genetic variability of introduced and local Spanish peach cultivars determined by SSR markers. *Tree Genetics & Genomes, 7*(2), 257-270.

Bousalem, M., Candresse, T., Quiot‐Douine, L., & Quiot, J. (1994). Comparison of three methods for assessing *plum pox virus* variability: further evidence for the existence of two major groups of isolates. *Journal of Phytopathology, 142*(2), 163-172.

Boyko, A., Zemp, F., Filkowski, J., & Kovalchuk, I. (2006). Double-strand break repair in plants is developmentally regulated. *Plant physiology, 141*(2), 488.

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 Biol, 11*(1), 24.

Braun, D.M., & Walker, J.C. (1996). Plant transmembrane receptors: new pieces in the signaling puzzle. *Trends in Biochemical Sciences, 21*(2), 70-73.

Brisson, N., Paszkowski, J., Penswick, J.R., Gronenborn, B., Potrykus, I., & Hohn, T. (1984). Expression of a bacterial gene in plants by using a viral vector. *nature, 310*, 511.

Burch-Smith, T.M., Anderson, J.C., Martin, G.B., & Dinesh-Kumar, S.P. (2004). Applications and advantages of virus-induced gene silencing for gene function studies in plants. *The Plant Journal, 39*(5), 734-746.

Byrne, D.H., Nikolic, A.N., & Burns, E.E. (1991). Variability in sugars, acids, firmness, and color characteristics of 12 peach genotypes. *Journal of the American Society for Horticultural Science, 116*(6), 1004-1006.

Byrne, D.H., Raseira, M.B., Bassi, D., Piagnani, M.C., Gasic, K., Reighard, G.L., Moreno, M.A., & Pérez, S. (2012). Peach. In *Fruit Breeding* (pp. 505-569): Springer.

Cabo, S., Carvalho, A., Martín, A., & Lima-Brito, J. (2014). Structural rearrangements detected in newly-formed hexaploid tritordeum after three sequential FISH experiments with repetitive DNA sequences. *Journal of genetics, 93*(1), 183-188.

Cambra, M., Asensio, M., Gorris, M.T., Pérez, E., Camarasa, E., Garciá, J.A., Moya, J.J., López-Abella, D., Vela, C., & Sanz, A. (1994). Detection of *plum pox potyvirus* using monoclonal antibodies to structural and non-structural proteins. *EPPO Bulletin, 24*(3), 569-577.

Cantín, C.M., Arús, P., & Eduardo, I. (2018). Identification of a new allele of the *Dw* gene causing brachytic dwarfing in peach. *BMC Research Notes, 11*(1), 386.

Cantín, C.M., Gogorcena, Y., & Moreno, M.Á. (2009). Analysis of phenotypic variation of sugar profile in different peach and nectarine [*Prunus persica* (L.) Batsch] breeding progenies. *Journal of the Science of Food and Agriculture, 89*(11), 1909-1917.

Cao, K., Wang, L., Zhu, G., Fang, W., Chen, C., & Zhao, P. (2011). Construction of a linkage map and identification of resistance gene analog markers for root-knot nematodes in wild peach, *Prunus kansuensis*. *Journal of the American Society for Horticultural Science, 136*(3), 190-197.

Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W., Cheng, S., Zeng, P., Chen, C., Wang, X., Xie, M., Zhong, X., Wang, X., Zhao, P., Bian, C., Zhu, Y., Zhang, J., Ma, G., Chen, C., Li, Y., Hao, F., Li, Y., Huang, G., Li, Y., Li, H., Guo, J., Xu, X., & Wang, J. (2014). Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. *Genome biology, 15*(7), 415.

Cao, K., Zhou, Z., Wang, Q., Guo, J., Zhao, P., Zhu, G., Fang, W., Chen, C., Wang, X., Wang, X., Tian, Z., & Wang, L. (2016). Genome-wide association study of 12 agronomic traits in peach. *Nature Communication, 7*, 13246.

Capecchi, M.R. (1989). Altering the genome by homologous recombination. *Science, 244*(4910), 1288.

Carbonell-Bejerano, P., Royo, C., Torres-Pérez, R., Grimplet, J., Fernandez, L., Franco-Zorrilla, J.M., Lijavetzky, D., Baroja, E., Martínez, J., García-Escudero, E., Ibáñez, J., & Martínez-Zapater, J.M. (2017). Catastrophic unbalanced genome rearrangements cause somatic loss of berry color in grapevine. *Plant physiology*.

Carrier, G., Le Cunff, L., Dereeper, A., Legrand, D., Sabot, F., Bouchez, O., Audeguin, L., Boursiquot, J.-M., & This, P. (2012). Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. *PLoS One, 7*(3), e32973.

Carrier, G., Le Cunff, L., Dereeper, A., Legrand, D., Sabot, F., Bouchez, O., Audeguin, L., Boursiquot, J.M., & This, P. (2012). Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. *PLoS One, 7*(3), e32973.

Carrington, J.C., & Dougherty, W.G. (1987). Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *Journal of virology, 61*(8), 2540.

Chang, C., & Stadler, R. (2001). Ethylene hormone receptor action in *Arabidopsis*. *BioEssays, 23*(7), 619-627.

Chaparro, J.X., Werner, D.J., O'Malley, D., & Sederoff, R.R. (1994). Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. *Theoretical and Applied Genetics, 87*(7), 805-815.

Chapman, S., Kavanagh, T., & Baulcombe, D. (1992). *Potato virus X* as a vector for gene expression in plants. *The Plant Journal, 2*(4), 549-557.

Chatelet, P., Laucou, V., Fernandez, L., Sreekantan, L., Lacombe, T., Martinez-Zapater, J.M., Thomas, M.R., & Torregrosa, L. (2007). Characterization of *Vitis vinifera* L. somatic

variants exhibiting abnormal flower development patterns. *Journal of Experimental Botany, 58*(15-16), 4107-4118.

Chen, Y., Mao, Y., Liu, H., Yu, F., Li, S., & Yin, T. (2014). Transcriptome analysis of differentially expressed genes relevant to variegation in peach flowers. *PLoS One, 9*(3), e90842.

Chin, S., Shaw, J., Haberle, R., Wen, J., & Potter, D. (2014). Diversification of almonds, peaches, plums and cherries–molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Molecular phylogenetics and evolution, 76*, 34-48.

Choi, I.-R., Stenger, D.C., Morris, T.J., & French, R. (2001). A plant virus vector for systemic expression of foreign genes in cereals. *The Plant Journal, 23*(4), 547-555.

Chow, C.N., Zheng, H.Q., Wu, N.Y., Chien, C.H., Huang, H.D., Lee, T.Y., Chiang-Hsieh.Y.F., Hou, P.F., Yang, T.Y., & Chang, W.C. (2015). PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic acids research, 44*(D1), D1154-D1160.

Chusreeaeom, K., Ariizumi, T., Asamizu, E., Okabe, Y., Shirasawa, K., & Ezura, H. (2014). A novel tomato mutant, Solanum lycopersicumelongated fruit1 (*Slelf1*), exhibits an elongated fruit shape caused by increased cell layers in the proximal region of the ovary. *Molecular Genetics and Genomics, 289*(3), 399-409.

Cirilli, M., Bassi, D., & Ciacciulli, A. (2016). Sugars in peach fruit: a breeding perspective. *Hortic Res, 3*, 15067.

Cirilli, M., Rossini, L., Geuna, F., Palmisano, F., Minafra, A., Castrignanò, T., Gattolin, S., Ciacciulli, A., Babini, A.R., Liverani, A., & Bassi, D. (2017). Genetic dissection of Sharka disease tolerance in peach (*P. persica* L. Batsch). *BMC Plant Biol, 17*(1), 192.

Clark, S.E., Running, M.P., & Meyerowitz, E.M. (1993). *CLAVATA1*, a regulator of meristem and flower development in *Arabidopsis*. *Development, 119*(2), 397.

Claverie, M., Bosselut, N., Lecouls, A.C., Voisin, R., Lafargue, B., Poizat, C., Kleinhentz, M., Laigret, F., Dirlewanger, E., & Esmenjaud, D. (2004a). Location of independent root-knot nematode resistance genes in plum and peach. *Theoretical and Applied Genetics, 108*(4), 765-773.

Cock, J.M., Vanoosthuyse, V., & Gaude, T. (2002). Receptor kinase signalling in plants and animals: distinct molecular systems with mechanistic similarities. *Current Opinion in Cell Biology, 14*(2), 230-236.

Coulson, A., Sulston, J., Brenner, S., & Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences, 83*(20), 7821-7825.

Cui, H., & Wang, A. (2016). An efficient viral vector for functional genomic studies of *Prunus* fruit trees and its induced resistance to *Plum pox virus* via silencing of a host factor gene. *Plant Biotechnology Journal, 15*(3), 344-356.

Dardick, C.D., Callahan, A.M., Chiozzotto, R., Schaffer, R.J., Piagnani, M.C., & Scorza, R. (2010). Stone formation in peach fruit exhibits spatial coordination of the lignin and flavonoid pathways and similarity to *Arabidopsis* dehiscence. *BMC Biology, 8*(1), 13.

De Smet, I., Voß, U., Jürgens, G., & Beeckman, T. (2009). Receptor-like kinases shape the plant. *Nature Cell Biology, 11*, 1166.

Decroocq, V., Foulongne, M., Lambert, P., Gall, O.L., Mantin, C., Pascal, T., Schurdi-Levraud, V., & Kervella, J. (2005). Analogues of virus resistance genes map to QTLs for resistance to sharka disease in *Prunus* davidiana. *Molecular Genetics and Genomics, 272*(6), 680-689.

Decroocq, V., Salvador, B., Sicard, O., Glasa, M., Cosson, P., Svanella-Dumas, L., Revers, F., García, J.A., & Candresse, T. (2009). The determinant of potyvirus ability to overcome the RTM resistance of *Arabidopsis thaliana* maps to the N-terminal region of the coat protein. *Molecular Plant-Microbe Interactions, 22*(10), 1302-1311.

Dermen, H. (1948). Chimeral apple sports: And their propagation through adventitious buds. *Journal of heredity, 39*(8), 235-242.

Dermen, H. (1953). Periclinal cytochimeras and origin of tissues in stem and leaf of peach. *American journal of botany, 40*(3), 154-168.

Dettori, M.T., Quarta, R., & Verde, I. (2001). A peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. *Genome, 44*(5), 783-790.

Díaz-Vivancos, P., Clemente-Moreno, M.J., Rubio, M., Olmos, E., García, J.A., Martínez-Gómez, P., & Hernández, J.A. (2008). Alteration in the chloroplastic metabolism leads to ROS accumulation in pea plants in response to *Plum pox viru*. *Journal of Experimental Botany, 59*(8), 2147-2160.

Dirlewanger, E., Cosson, P., Boudehri, K., Renaud, C., Capdeville, G., Tauzin, Y., Laigret, F., & Moing, A. (2006). Development of a second-generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. *Tree Genetics & Genomes, 3*(1), 1-13.

Dirlewanger, E., Cosson, P., Howad, W., Capdeville, G., Bosselut, N., Claverie, M., Voisin, R., Poizat, C., Lafargue, B., Baron, O., Laigret, F., Kleinhentz, M., Arús, P., & Esmenjaud, D. (2004b). Microsatellite genetic linkage maps of myrobalan plum and an almond-peach hybrid—location of root-knot nematode resistance genes. *Theoretical and Applied Genetics, 109*(4), 827-838.

Dirlewanger, E., Graziano, E., Joobeur, T., Garriga-Calderé, F., Cosson, P., Howad, W., & Arús, P. (2004a). Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proceedings of the National Academy of Sciences, 101*(26), 9891-9896.

Dirlewanger, E., Pronier, V., Parvery, C., Rothan, C., Guye, A., & Monet, R. (1998). Genetic linkage map of peach [*Prunus persica* (L.) Batsch] using morphological and molecular markers. *Theoretical and Applied Genetics, 97*(5-6), 888-895.

Donoso, J.M., Eduardo, I., Picañol, R., Batlle, I., Howad, W., Aranzana, M.J., & Arús, P. (2015). High-density mapping suggests cytoplasmic male sterility with two restorer genes in almond × peach progenies. *Hortic Res, 2*, 15016.

Donoso, J.M., Picañol, R., Serra, O., Howad, W., Alegre, S., Arús, P., & Eduardo, I. (2016). Exploring almond genetic variability useful for peach improvement: mapping major

genes and QTLs in two interspecific almond× peach populations. *Molecular Breeding, 36*(2), 16.

Doudna, J.A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science, 346*(6213).

Doyle, J.J. (1990). Isolation of plant DNA from fresh tissue. *Focus, 12*, 13-15.

Ebihara, K., Kiriyama, S., & Manabe, M. (1979). Cholesterol-lowering activity of various natural pectins and synthetic pectin-derivatives with different physicochemical properties. *NUTRITION REPORTS INTERNATIONAL, 20*(4), 519-526.

Eduardo, I., Chietera, G., Bassi, D., Rossini, L., & Vecchietti, A. (2010). Identification of key odor volatile compounds in the essential oil of nine peach accessions. *Journal of the Science of Food and Agriculture, 90*(7), 1146-1154.

Eduardo, I., López-Girona, E., Batlle, I., Reig, G., Iglesias, I., Howad, W., Arús, P., & Aranzana, M.J. (2014). Development of diagnostic markers for selection of the subacid trait in peach. *Tree Genetics & Genomes, 10*(6), 1695-1709.

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]. *Theoretical and Applied Genetics, 105*(1), 145-159.

Falchi, R., Vendramin, E., Zanon, L., Scalabrin, S., Cipriani, G., Verde, I., Vizzotto, G., & Morgante, M. (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *Plant J, 76*(2), 175-187.

Falchi, R., Vendramin, E., Zanon, L., Scalabrin, S., Cipriani, G., Verde, I., Vizzotto, G., & Morgante, M. (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *The Plant Journal, 76*(2), 175-187.

Fan, S., Bielenberg, D.G., Zhebentyayeva, T.N., Reighard, G.L., Okie, W.R., Holland, D., & Abbott, A.G. (2010). Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). *New Phytologist, 185*(4), 917-930.

FAOSTAT. (2016). Production/Area harvested/Yield of peaches and nectarines in 2016. from Food and Agricultural Organization, Statistics Division http://www.fao.org/faostat

Faust, M., & Timon, B. (1995). Origin and dissemination of peach. *Horticulture Reviews, 17*, 331-379.

Fernandez, L., Torregrosa, L., Segura, V., Bouquet, A., & Martinez-Zapater, J.M. (2010). Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. *The Plant Journal, 61*(4), 545-557.

Fischer, I., Diévart, A., Droc, G., Dufayard, J.-F., & Chantret, N. (2016). Evolutionary dynamics of the leucine-rich repeat receptor-like kinase (LRR-RLK) subfamily in angiosperms. *Plant physiology, 170*(3), 1595.

Fontes, E.P.B., Santos, A.A., Luz, D.F., Waclawovsky, A.J., & Chory, J. (2004). The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. *Genes & development, 18*(20), 2545-2556.

Foolad, M.R., Arulsekar, S., Becerra, V., & Bliss, F.A. (1995). A genetic map of *Prunus* based on an interspecific cross between peach and almond. *Theoretical and Applied Genetics, 91*(2), 262-269.

Foster, T.M., & Aranzana, M.J. (2018). Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Hortic Res, 5*(1), 44.

Foulongne, M., Pascal, T., Pfeiffer, F., & Kervella, J. (2003). QTLs for powdery mildew resistance in peach ×*Prunus* davidiana crosses: consistency across generations and environments. *Molecular Breeding, 12*(1), 33-50.

Fu, H., & Luan, S. (1998). AtKUP1: A dual-affinity K$^+$ transporter from *Arabidopsis*. *Plant Cell, 10*(1), 63.

García, J.A., Glasa, M., Cambra, M., & Candresse, T. (2013). *Plum pox virus* and sharka: a model potyvirus and a major disease. *Molecular Plant Pathology, 15*(3), 226-241.

Garg, R., Patel, R.K., Tyagi, A.K., & Jain, M. (2011). De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA research, 18*(1), 53-63.

Germana, M.A. (2006). Doubled haploid production in fruit crops. *Plant Cell, Tissue and Organ Culture, 86*(2), 131.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods, 6*(5), 343-345.

Gillen, A.M., & Bliss, F.A. (2005). Identification and mapping of markers linked to the *Mi* gene for root-knot nematode resistance in peach. *Journal of the American Society for Horticultural Science, 130*(1), 24-33.

Giovannoni, J.J., Noensie, E.N., Ruezinsky, D.M., Lu, X., Tracy, S.L., Ganal, M.W., Martin, G.B., Pillen, K., Albert, K., & Tankslev, S.D. (1995). Molecular genetic analysis of *theripening-inhibitor* and *non-ripening* loci of tomato: A first step in genetic map-based cloning of fruit ripening genes. *Molecular and General Genetics MGG, 248*(2), 195-206.

Glasa, M., Palkovics, L., Komínek, P., Labonne, G., Pittnerová, S., Kúdela, O., Candresse, T., & Šubr, Z. (2004). Geographically and temporally distant natural recombinant isolates of *Plum pox virus* (PPV) are genetically very similar and form a unique PPV subgroup. *Journal of General Virology, 85*(9), 2671-2681.

Gómez-Gómez, L., & Boller, T. (2000). *FLS2*: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Molecular cell, 5*(6), 1003-1011.

Green, E.D. (2001). Strategies for the systematic sequencing of complex genomes. *Nature Reviews Genetics, 2*(8), 573.

Guo, H.S., López-Moya, J.J., & García, J.A. (1998). Susceptibility to recombination rearrangements of a chimeric plum pox potyvirus genome after insertion of a foreign gene. *Virus research, 57*(2), 183-195.

Guo, S., Iqbal, S., Ma, R., Song, J., Yu, M., & Gao, Z. (2018). High-density genetic map construction and quantitative trait loci analysis of the stony hard phenotype in peach based on restriction-site associated DNA sequencing. *BMC Genomics, 19*(1), 612.

Hahn, S., Buratowski, S., Sharp, P.A., & Guarente, L. (1989). Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proceedings of the National Academy of Sciences, 86*(15), 5718-5722.

Hauser, M.-T., Harr, B., & Schlötterer, C. (2001). Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: Molecular analysis of the candidate gene *GLABROUS1*. *Molecular Biology and Evolution, 18*(9), 1754-1763.

Hawkins, C., Caruana, J., Schiksnis, E., & Liu, Z.C. (2016). Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. *Sci Rep, 6*, 29017.

He, K., Gou, X., Yuan, T., Lin, H., Asami, T., Yoshida, S., Russell, S.D., & Li, J. (2007). *BAK1* and *BKK1* regulate brassinosteroid-dependent growth and brassinosteroid-independent cell-death pathways. *Current Biology, 17*(13), 1109-1115.

Hedrick, U.P. (1950). *A history of Horticulture in America to 1860*: Oxford University Press, New York, and Geoffrey Cumberlege, London.

Hedrick, U.P., Howe, G.H., Taylor, O.M., & Tubergen, C.B. (1917). *The peaches of New York* (Vol. 2): JB Lyon Company, printers.

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *nature, 329*(6136), 219-222.

Hervé, C., Dabos, P., Galaud, J.P., Rougé, P., & Lescure, B. (1996). Characterization of an *Arabidopsis thaliana* gene that defines a new class of putative plant receptor kinases with an extracellular lectin-like domain. *Journal of Molecular Biology, 258*(5), 778-788.

Hollender, C.A., Hadiarto, T., Srinivasan, C., Scorza, R., & Dardick, C. (2016). A brachytic dwarfism trait (*dw*) in peach trees is caused by a nonsense mutation within the gibberellic acid receptor *PpeGID1c*. *New Phytologist, 210*(1), 227-239.

Hollings, M., & Brunt, A.A. (1981). Potyviruses. *Handbook of Plant Virus Infections and Comparative Diagnosis*, 731-807.

Horn, R., Lecouls, A., Callahan, A., Dandekar, A., Garay, L., McCord, P., Howad, W., Chan, H., Verde, I., & Main, D. (2005). Candidate gene database and transcript map for peach, a model species for fruit trees. *Theoretical and Applied Genetics, 110*(8), 1419-1428.

Hoskins, R.A., Nelson, C.R., Berman, B.P., Laverty, T.R., George, R.A., Ciesiolka, L., Naeemuddin, M., Arenson, A.D., Durbin, J., & David, R.G. (2000). A BAC-based physical map of the major autosomes of *Drosophila* melanogaster. *Science, 287*(5461), 2271-2274.

Howad, W., Yamamoto, T., Dirlewanger, E., Testolin, R., Cosson, P., Cipriani, G., Monforte, A.J., Georgi, L., Abbott, A.G., & Arus, P. (2005). Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. *Genetics*.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A.F., Selengut, J.D., Sigrist, C.J.A., Thimma, M., Thomas, P.D., Valentin, F., Wilson, D., Wu, C.H., & Yeats, C. (2009). InterPro: the integrative protein signature database. *Nucleic acids research, 37*(suppl_1), D211-D215.

Iglesias, i.C.I. (2018). Tendencias e innovación, en las principales especies leñosas de fruta dulce en España. *Vida rural*(448), 20-28.

Islam, M.Q., & Levan, G. (1987). A new fixation procedure for improved quality G‑bands in routine cytogenetic work. *Hereditas, 107*(1), 127-130.

Jaenisch, R., & Mintz, B. (1974). Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proceedings of the National Academy of Sciences, 71*(4), 1250.

Janick, J., & Moore, J.N. (1996). *Fruit breeding, tree and tropical fruits* (Vol. 1): John Wiley & Sons.

Jáuregui, B., De Vicente, M.C., Messeguer, R., Felipe, A., Bonnet, A., Salesses, G., & Arús, P. (2001). A reciprocal translocation between 'Garfi' almond and 'Nemared' peach. *Theoretical and Applied Genetics, 102*(8), 1169-1176.

Jia, H., Guo, J., Qin, L., & Shen, Y. (2015). Virus-induced *PpCHLH* gene silencing in peach leaves (*Prunus persica*). *The Journal of Horticultural Science and Biotechnology, 85*(6), 528-532.

Jinn, T.-L., Stone, J.M., & Walker, J.C. (2000). *HAESA*, an *Arabidopsis* leucine-rich repeat receptor kinase, controls floral organ abscission. *Genes & development, 14*(1), 108-117.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res, 36*(Web Server issue), W5-9.

Joobeur, T., Periam, N., Vicente, M.C.d., King, G.J., & Arús, P. (2000). Development of a second generation linkage map for almond using RAPD and SSR markers. *Genome, 43*(4), 649-655.

Joobeur, T., Viruel, M.A., de Vicente, M.C., Jauregui, B., Ballester, J., Dettori, M.T., Verde, I., Truco, M.J., Messeguer, R., & Batlle, I. (1998). Construction of a saturated linkage map for *Prunus* using an almond × peach F$_2$ progeny. *Theoretical and Applied Genetics, 97*(7), 1034-1041.

Jung, S., Ficklin, S.P., Lee, T., Cheng, C.-H., Blenda, A., Zheng, P., Yu, J., Bombarely, A., Cho, I., Ru, S., Evans, K., Peace, C., Abbott, A.G., Mueller, L.A., Olmstead, M.A., & Main, D. (2014). The Genome Database for Rosaceae (GDR): year 10 update. *Nucleic acids research, 42*(D1), D1237-D1244.

Jung, S., Jiwan, D., Cho, I., Lee, T., Abbott, A., Sosinski, B., & Main, D. (2009). Synteny of *Prunus* and other model plant species. *BMC Genomics, 10*(1), 76.

Jung, S., Main, D., Staton, M., Cho, I., Zhebentyayeva, T., Arús, P., & Abbott, A. (2006). Synteny conservation between the *Prunus* genome and both the present and ancestral *Arabidopsis* genomes. *BMC Genomics, 7*(1), 81.

Kayes, J.M., & Clark, S.E. (1998). *CLAVATA2*, a regulator of meristem and organ development in *Arabidopsis*. *Development, 125*(19), 3843-3851.

Keefe, C., McDevitt, M.A., & Maciejewski, J.P. (2010). Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood*.

Kim, Y.G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to *Fok I* cleavage domain. *Proceedings of the National Academy of Sciences, 93*(3), 1156.

Kirov, I., Divashuk, M., Van-Laere, K., Soloviev, A., & Khrustaleva, L. (2014). An easy SteamDrop method for high quality plant chromosome preparation. *Molecular Cytogenetics*.

Kobe, B., & Deisenhofer, J. (1994). The leucine-rich repeat: a versatile binding motif. *Trends in Biochemical Sciences, 19*(10), 415-421.

Kobe, B., & Deisenhofer, J. (1995). Proteins with leucine-rich repeats. *Current Opinion in Structural Biology, 5*(3), 409-416.

Kondrashov, F.A., & Koonin, E.V. (2004). A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends in Genetics, 20*(7), 287-290.

LaFave, M.C., & Sekelsky, J. (2009). Mitotic Recombination: Why? When? How? Where? *PLoS Genet, 5*(3), e1000411.

Lambert, P., Campoy, J.A., Pacheco, I., Mauroux, J.-B., Da Silva Linge, C., Micheletti, D., Bassi, D., Rossini, L., Dirlewanger, E., Pascal, T., Troggio, M., Aranzana, M.J., Patocchi, A., & Arús, P. (2016). Identifying SNP markers tightly associated with six major genes in peach [*Prunus persica* (L.) Batsch] using a high-density SNP array with an objective of marker-assisted selection (MAS). *Tree Genetics & Genomes, 12*(6).

Lambert, P., & Pascal, T. (2011). Mapping *Rm2* gene conferring resistance to the green peach aphid (*Myzus persicae* Sulzer) in the peach cultivar "Rubira®". *Tree Genetics & Genomes, 7*(5), 1057-1068.

Lansac, M., Eyquard, J.P., Salvador, B., Garcia, J.A., Le Gall, O., Decroocq, V., & Schurdi-Levraud Escalettes, V. (2005). Application of GFP-tagged *Plum pox virus* to study *Prunus*–PPV interactions at the whole plant and cellular levels. *Journal of Virological Methods, 129*(2), 125-133.

Layne, D.R., & Bassi, D. (2008). *The peach: botany, production and uses*: CABI.

Lee, T.I., & Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annu Rev Genet, 34*(1), 77-137.

Li, H. (1983). The domestication of plants in China: ecogeographical considerations. *The origins of Chinese civilization*, 21-63.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, J., Li, F., Qian, M., Han, M., Liu, H., Zhang, D., Ma, J., & Zhao, C. (2017). Characteristics and regulatory pathway of the *PrupeSEP1 SEPALLATA* gene during ripening and softening in peach fruits. *Plant Sci, 257*, 63-73.

Li, J., & Tax, F.E. (2013). Receptor-like kinases: key regulators of plant development and defense. *J Integr Plant Biol, 55*(12), 1184-1187.

Li, S. (2013). *The study of peach trees*. Beijing: China Agriculture Press.

Li, X., Jiang, J., Zhang, L., Yu, Y., Ye, Z., Wang, X., Zhou, J., Chai, M., Zhang, H., & Arús, P. (2015). Identification of volatile and softening-related genes using digital gene expression profiles in melting peach. *Tree Genetics & Genomes, 11*(4), 71.

Li, X., Meng, X., Jia, H., Yu, M.-i., Ma, R., Wang, L., Cao, K., Shen, Z., Niu, L., Tian, J., Chen, M., Xie, M., Arus, P., Gao, Z., & Aranzana, M.J. (2013). Peach genetic resources: diversity, population structure and linkage disequilibrium. *BMC Genetics, 14*(1), 84.

Li, Z. (1984). Peach germplasm and breeding in China. *Horticulture Science*.

Lijavetzky, D., Ruiz-García, L., Cabezas, J.A., De Andrés, M.T., Bravo, G., Ibáñez, A., Carreño, J., Cabello, F., Ibáñez, J., & Martínez-Zapater, J.M. (2006). Molecular genetics of berry colour variation in table grape. *Molecular Genetics and Genomics, 276*(5), 427-435.

Lindbo, J.A. (2007). TRBO: A high-efficiency *tobacco mosaic virus* RNA-based overexpression vector. *Plant physiology, 145*(4), 1232.

Liu, J., Van Eck, J., Cong, B., & Tanksley, S.D. (2002). A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proceedings of the National Academy of Sciences, 99*(20), 13302.

Llácer, G., Alonso, J.M., Rubio-Cabetas, M.J., Batlle, I., Iglesias, I., Vargas, F.J., García-Brunton, J., & Bardenes, M.L. (2009). Peach industry in Spain. *Journal of the American Pomological Society, 63*(3), 128.

López-Girona, E., Zhang, Y., Eduardo, I., Mora, J.R.H., Alexiou, K.G., Arus, P., & Aranzana, M.J. (2017). A deletion affecting an LRR-RLK gene co-segregates with the fruit flat shape trait in peach. *Sci Rep, 7*(1), 6714.

López-Moya, J.J., Fernández-Fernández, M.a.R., Cambra, M., & García, J.A. (2000). Biotechnological aspects of *plum pox virus*. *J Biotechnol, 76*(2), 121-136.

López-Moya, J.J., García, J.A. (2000). Construction of a stable and highly infectious intron-containing cDNA clone of *Plum pox potyvirus* and its use to infect plants by particle bombardment. *Virus research, 68*(2), 99-107.

López-Moya, J.J., Valli, A., & García, J.A. (2009). Potyviridae. *eLS,* DOI:10.1002/9780470015902.a0000755.pub2.

López Girona, E. (2014). *Genetic architecture of agronomic traits in peach [Prunus persica (L.) Batsch]: subacid, flat shape and nectarine.* (PhD), Universitat Autònoma de Barcelona, Barcelona.

Lu, Z., Sosinski, B., Reighard, G.L., Baird, W.V., & Abbott, A.G. (1998). Construction of a genetic linkage map and identification of AFLP markers for resistance to root-knot nematodes in peach rootstocks. *Genome, 41*(2), 199-207.

Lu, Z., Sossey-Alaoui, K., Reighard, G.L., Baird, W.V., & Abbott, A.G. (1999). Development and characterization of a codominant marker linked to root-knot nematode resistance, and its application to peach rootstock breeding. *Theoretical and Applied Genetics, 99*(1), 115-122.

Ma, Y., Islam-Faridi, M.N., Crane, C.F., Stelly, D.M., Price, H.J., & Byrne, D.H. (1996). A new procedure to prepare slides of metaphase chromosomes of roses. *HortScience, 31*(5), 855-857.

Maiss, E., Timpe, U., Brisske, A., Jelkmann, W., Casper, R., Himmler, G., Mattanovich, D., & Katinger, H.W.D. (1989). The complete nucleotide sequence of *plum pox virus* RNA. *Journal of General Virology, 70*(3), 513-524.

Mandel, T., Moreau, F., Kutsher, Y., Fletcher, J.C., Carles, C.C., & Williams, L.E. (2014). The *ERECTA* receptor kinase regulates *Arabidopsis* shoot apical meristem size, phyllotaxy and floral meristem identity. *Development, 141*(4), 830.

Manova, V., & Gruszka, D. (2015). DNA damage and repair in plants - from models to crops. *Front Plant Sci, 6*, 885.

Maquilan, M.A.D, Olmstead, M.A., Olmstead, J.W., Dickson, D.W., & Chaparro, J.X. (2018). Genetic analyses of resistance to the peach root-knot nematode (*Meloidogyne floridensis*) using microsatellite markers. *Tree Genetics & Genomes, 14*(4), 47.

Marandel, G., Pascal, T., Candresse, T., & Decroocq, V. (2009). Quantitative resistance to *Plum pox virus* in *Prunus* davidiana P1908 linked to components of the eukaryotic translation initiation complex. *Plant pathology, 58*(3), 425-435.

Marra, M., Kucaba, T., Sekhon, M., Hillier, L., Martienssen, R., Chinwalla, A., Fedele, J., Grover, H., Gund, C., & McCombie, W.R. (1999). A map for sequence analysis of the *Arabidopsis thaliana* genome. *Nature Genetics, 22*(3), 265.

Marroni, F., Scaglione, D., Pinosio, S., Policriti, A., Miculan, M., Di Gaspero, G., & Morgante, M. (2017). Reduction of heterozygosity (ROH) as a method to detect mosaic structural variation. *Plant Biotechnol J, 15*(7), 791-793.

Marroni, F., Scaglione, D., Pinosio, S., Policriti, A., Miculan, M., Di Gaspero, G., & Morgante, M. (2017). Reduction of heterozygosity (ROH) as a method to detect mosaic structural variation. *Plant Biotechnology Journal, 15*(7), 791-793.

Martínez-García, P.J., Parfitt, D.E., Ogundiwin, E.A., Fass, J., Chan, H.M., Ahmad, R., Lurie, S., Dandekar, A., Gradziel, T.M., & Crisosto, C.H. (2013). High density SNP mapping and QTL analysis for fruit quality characteristics in peach (Prunus persica L.). *Tree Genetics & Genomes, 9*(1), 19-36.

Martínez-Gomez, P., Audergon, J.M., & Dicenta, F. (2000). Efficiency of inoculation of peach GF305 seedlings with *Plum pox virus* by different methods. *Acta virologica, 44*(6), 329-333.

Matsubayashi, Y., Ogawa, M., Morita, A., & Sakagami, Y. (2002). An LRR receptor kinase involved in perception of a peptide plant hormone, phytosulfokine. *Science, 296*(5572), 1470.

Matsushima, N., & Miyashita, H. (2012). Leucine-rich repeat (LRR) domains containing intervening motifs in plants. *Biomolecules, 2*(2).

McCarty, D.R., & Chory, J. (2000). Conservation and innovation in plant signaling pathways. *Cell, 103*(2), 201-209.

Mei, Y., & Whitham, S.A. (2018). Virus-induced gene silencing in maize with a foxtail mosaic virus vector. In Lagrimini, L. M. (Ed.), *Maize: Methods and Protocols* (10.1007/978-1-4939-7315-6_7pp. 129-139). New York, NY: Springer New York.

Micheletti, D., Dettori, M.T., Micali, S., Aramini, V., Pacheco, I., Da Silva Linge, C., Foschi, S., Banchi, E., Barreneche, T., Quilot-Turion, B., Lambert, P., Pascal, T., Iglesias, I., Carbo, J., Wang, L.R., Ma, R.J., Li, X.W., Gao, Z.S., Nazzicari, N., Troggio, M., Bassi, D., Rossini, L., Verde, I., Laurens, F., Arus, P., & Aranzana, M.J. (2015). Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. *PLoS One, 10*(9), e0136803.

Milne, I., Shaw, P., Stephen, G., Bayer, M., Cardle, L., Thomas, W.T.B., Flavell, A.J., & Marshall, D. (2010). Flapjack—graphical genotype visualization. *Bioinformatics, 26*(24), 3133-3134.

Moon, S., Jung, K.-H., Lee, D.-E., Lee, D.-Y., Lee, J., An, K., Kang, H.-G., & An, G. (2006). The rice *FON1* gene controls vegetative and reproductive development by regulating shoot apical meristem size. *Molecules & Cells (Springer Science & Business Media BV), 21*(1).

Moynahan, M.E., & Jasin, M. (1997). Loss of heterozygosity induced by a chromosomal double-strand break. *Proceedings of the National Academy of Sciences, 94*(17), 8988-8993.

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., Carretero, Y., Le Paslier, M.-C., Delalande, C., Bouzayen, M., Brunel, D., & Causse, M. (2011). Increase in tomato locule number is controlled by two SNPs located near *WUSCHEL*. *Plant physiology*.

Myers, S.C., Okie, W.R., & Lightner, G. (1989). The 'Elberta' peach. *Fruit varieties journal (USA)*.

Narzisi, G., & Mishra, B. (2011). Comparing De Novo Genome Assembly: The Long and Short of It. *PLoS One, 6*(4), e19175.

O'Rourke, J.A. (2014). Genetic and Physical Map Correlation. *eLS,* doi:10.1002/9780470015902.a0000819.pub3

10.1002/9780470015902.a0000819.pub3.

Ogundiwin, A., Peace, C.P., Gradziel, T.M., Parfitt, D.E., Bliss, F.A., & Crisosto, C.H. (2009). A fruit quality gene map of *Prunus*. *BMC Genomics, 10*(1), 587.

Okie, W.R. (1998). *Handbook of Peach and Nectarine Varieties*.

Orphanides, G., Lagrange, T., & Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes & development, 10*(21), 2657-2683.

Osakabe, Y., Maruyama, K., Seki, M., Satou, M., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2005). Leucine-rich repeat receptor-like kinase 1 is a key membrane-bound regulator of abscisic acid early signaling in *Arabidopsis*. *Plant Cell, 17*(4), 1105.

Osorio, S., Alba, R., Damasceno, C.M.B., López-Casado, G., Lohse, M., Zanor, M.I., Tohge, T., Usadel, B., Rose, J.K.C., Fei, Z., Giovannoni, J.J., & Fernie, A.R. (2011). Systems biology of tomato fruit development: combined transcript, protein and metabolite analysis of tomato transcription factor (*nor*, *rin*) and ethylene receptor (*Nr*) mutants reveals novel regulatory interactions. *Plant physiology*.

Otto, D., Petersen, R., Brauksiepe, B., Braun, P., & Schmidt, E.R. (2013). The columnar mutation ("*Co* gene") of apple (*Malus × domestica*) is associated with an integration of a Gypsy-like retrotransposon. *Molecular Breeding, 33*(4), 863-880.

Padilla, I.M.G., Golis, A., Gentile, A., Damiano, C., & Scorza, R. (2006). Evaluation of transformation in peach *Prunus persica* explants using green fluorescent protein (GFP) and beta-glucuronidase (GUS) reporter genes. *Plant Cell, Tissue and Organ Culture, 84*(3), 309-314.

Pan, Y., Liang, X., Gao, M., Liu, H., Meng, H., Weng, Y., & Cheng, Z. (2017). Round fruit shape in WI7239 cucumber is controlled by two interacting quantitative trait loci with one putatively encoding a tomato *SUN* homolog. *Theoretical and Applied Genetics, 130*(3), 573-586.

Paran, I., & van der Knaap, E. (2007). Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *Journal of Experimental Botany, 58*(14), 3841-3852.

Pardo, B., Gómez-González, B., & Aguilera, A. (2009). DNA Repair in Mammalian Cells. *Cellular and Molecular Life Sciences, 66*(6), 1039-1056.

Pascal, T., Aberlenc, R., Confolent, C., Hoerter, M., Lecerf, E., Tuéro, C., & Lambert, P. (2017). Mapping of new resistance (*Vr2, Rm1*) and ornamental (*Di2, pl*) Mendelian trait loci in peach. *Euphytica, 213*(6), 132.

Pascal, T., Pfeiffer, F., & Kervella, J. (2010). Powdery mildew resistance in the peach cultivar Pamirskij 5 is genetically linked with the *Gr* gene for leaf color. *HortScience, 45*(1), 150-152.

Pelsy, F., Dumas, V., Bevilacqua, L., Hocquigny, S., & Merdinoglu, D. (2015). Chromosome replacement and deletion lead to clonal polymorphism of berry color in grapevine. *PLoS Genet, 11*(4), e1005081.

Pérez-Clemente, R.M., Pérez-Sanjuán, A., García-Férriz, L., Beltrán, J.-P., & Cañas, L.A. (2005). Transgenic peach plants (*Prunus persica* L.) produced by genetic transformation of embryo sections using the green fluorescent protein (GFP) as an in vivo marker. *Molecular Breeding, 14*(4), 419-427.

Picañol, R., Eduardo, I., Aranzana, M.J., Howad, W., Batlle, I., Iglesias, I., Alonso, J.M., & Arús, P. (2013). Combining linkage and association mapping to search for markers linked to the flat fruit character in peach. *Euphytica, 190*(2), 279-288.

Pirona, R., Eduardo, I., Pacheco, I., Da Silva Linge, C., Miculan, M., Verde, I., Tartarini, S., Dondini, L., Pea, G., Bassi, D., & Rossini, L. (2013). Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. *BMC Plant Biol, 13*(1), 166.

Pirovano, A. (1953). *New Italian peaches*: Roma.

Porebski, S., Bailey, L.G., & Baum, B.R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant molecular biology reporter, 15*(1), 8-15.

Postel, S., Küfner, I., Beuter, C., Mazzotta, S., Schwedt, A., Borlotti, A., Halter, T., Kemmerling, B., & Nürnberger, T. (2010). The multifunctional leucine-rich repeat receptor kinase *BAK1* is implicated in *Arabidopsis* development and immunity. *European Journal of Cell Biology, 89*(2), 169-174.

Potter, D., Eriksson, T., Evans, R.C., Oh, S., Smedmark, J.E.E., Morgan, D.R., Kerr, M., Robertson, K.R., Arsenault, M., & Dickinson, T.A. (2007). Phylogeny and classification of Rosaceae. *Plant systematics and evolution, 266*(1-2), 5-43.

Proost, S., Van Bel, M., Vaneechoutte, D., Van de Peer, Y., Inzé, D., Mueller-Roeber, B., & Vandepoele, K. (2015). PLAZA 3.0: an access point for plant comparative genomics. *Nucleic acids research, 43*(D1), D974-D981.

Quarta, R., Dettori, M.T., Sartori, A., & Verde, I. (2000). *Genetic linkage map and QTL analysis in peach*.

Quarta, R., Dettori, M.T., Verde, I., Gentile, A., & Broda, Z. (1998). *Genetic analysis of agronomic traits and genetic linkage mapping in a BC1 peach population using RFLPs and RAPDs*.

Quilot, B., Wu, B.H., Kervella, J., Génard, M., Foulongne, M., & Moreau, K. (2004). QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana*. *Theoretical and Applied Genetics, 109*(4), 884-897.

Rajapakse, S., Belthoff, L.E., He, G., Estager, A.E., Scorza, R., Verde, I., Ballard, R.E., Baird, W.V., Callahan, A., Monet, R., & Abbott, A.G. (1995). Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. *Theoretical and Applied Genetics, 90*(3), 503-510.

Ratcliff, F., Martin-Hernandez, A.M., & Baulcombe, D.C. (2008). Technical Advance: *Tobacco rattle virus* as a vector for analysis of gene function by silencing. *The Plant Journal, 25*(2), 237-245.

Rehder, A. (1940). *Manual of cultivated trees and shrubs*. New York: The Macmillan Company.

Rey, M.-D., Moore, G., & Martín, A.C. (2018). Identification and comparison of individual chromosomes of three accessions of *Hordeum chilense, Hordeum vulgare*, and *Triticum aestivum* by *FISH*. *Genome, 61*(6), 387-396.

Riechmann, J., Laín, S., & Garcia, J.A. (1990). Infectious *in vitro* transcripts from a *plum pox potyvirus* cDNA clone. *Virology, 177*(2), 710-716.

Roach, F.A. (1985). *Cultivated fruits of Britain: their origin and history*: Basil Blackwell Publisher Ltd.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., & Mesirov, J.P. (2011). Integrative genomics viewer. *Nat Biotechnol, 29*, 24.

Rodgers-Melnick, E., Mane, S.P., Dharmawardhana, P., Slavov, G.T., Crasta, O.R., Strauss, S.H., Brunner, A.M., & Difazio, S.P. (2012). Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. *Genome Res, 22*(1), 95-105.

Rodríguez, G.R., Kim, H.J., & van der Knaap, E. (2013). Mapping of two suppressors of OVATE (*sov*) loci in tomato. *Heredity, 111*, 256.

Rodriguez, G.R., Munos, S., Anderson, C., Sim, S.C., Michel, A., Causse, M., Gardener, B.B., Francis, D., & van der Knaap, E. (2011). Distribution of *SUN*, *OVATE*, *LC*, and *FAS* in the tomato germplasm and the relationship to fruit shape diversity. *Plant physiology, 156*(1), 275-285.

Rozen, S., & Skaletsky, H. (1999). Primer3 on the WWW for general users and for biologist programmers. In Misener, S. & Krawetz, S. A. (Eds.), *Bioinformatics Methods and Protocols* (10.1385/1-59259-192-2:365pp. 365-386). Totowa, NJ: Humana Press.

Rubio, M., Pascal, T., Bachellez, A., & Lambert, P. (2010). Quantitative trait loci analysis of *Plum pox virus* resistance in *Prunus davidiana* P1908: new insights on the organization of genomic resistance regions. *Tree Genetics & Genomes, 6*(2), 291-304.

Rubio, M., Rodriguez-Moreno, L., Ballester, A.R., de Moura, M.C., Bonghi, C., Candresse, T., & Martinez-Gomez, P. (2015). Analysis of gene expression changes in peach leaves in response to *Plum pox virus* infection using RNA-Seq. *Mol Plant Pathol, 16*(2), 164-176.

Ryland, G.L., Doyle, M.A., Goode, D., Boyle, S.E., Choong, D.Y.H., Rowley, S.M., Li, J., Bowtell, D.D.L., Tothill, R.W., Campbell, I.G., Gorringe, K.L., & Australian Ovarian Cancer Study, G. (2015). Loss of heterozygosity: what is it good for? *BMC Medical Genomics, 8*(1), 45.

Ryu, J., & Hartin, R.J. (1990). Quick transformation in Salmonella typhimurium LT2. *BioTechniques, 8*(1), 43-45.

Sablowski, R. (2007). Flowering and determinacy in *Arabidopsis*. *Journal of Experimental Botany, 58*(5), 899-907.

Sasidharan, R., & Gerstein, M. (2008). Protein fossils live on as RNA. *nature, 453*, 729.

Schäffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., & Altschul, S.F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic acids research, 29*(14), 2994-3005.

Scholthof, H.B., Scholthof, K.-B.G., & Jackson, A.O. (1996). Plant virus gene vectors for transient expression of foreign proteins in plants. *Annual review of Phytopathology, 34*(1), 299-323.

Scorza, R., Mehlenbacher, S.A., & Lightner, G.W. (1985). Inbreeding and coancestry of freestone peach cultivars of the eastern United States and implications for peach germplasm improvement. *Journal of the American Society for Horticultural Science (USA)*.

Scorza, R., Morgens, P.H., Cordts, J.M., Mante, S., & Callahan, A.M. (1990). Agrobacterium-mediated transformation of peach (Prunus persica L. batsch) leaf segments, immature embryos, and long-term embryogenic callus. *In Vitro Cellular &Developmental Biology, 26*(8), 829-834.

Seidman, J.G., & Seidman, C. (2002). Transcription factor haploinsufficiency: when half a loaf is not enough. *The Journal of Clinical Investigation, 109*(4), 451-455.

Serra, O., Donoso, J.M., Picañol, R., Batlle, I., Howad, W., Eduardo, I., & Arús, P. (2016). Marker-assisted introgression (MAI) of almond genes into the peach background: a fast method to mine and integrate novel variation from exotic sources in long intergeneration species. *Tree Genetics & Genomes, 12*(5), 96.

Shah, D.M., Tumer, N.E., Fischhoff, D.A., Horsch, R.B., Rogers, S.G., Fraley, R.T., & Jaworski, E.G. (1987). The introduction and expression of foreign genes in plants. *Biotechnology and Genetic Engineering Reviews, 5*(1), 81-106.

Shahabuddin, M., Shaw, J.G., & Rhoads, R.E. (1988). Mapping of the tobacco vein mottling virus VPg cistron. *Virology, 163*(2), 635-637.

Shahmuradov, I.A., Umarov, R.K., & Solovyev, V.V. (2017). TSSPlant: a new tool for prediction of plant Pol II promoters. *Nucleic acids research, 45*(8), e65-e65.

Shiu, S.-H., & Bleecker, A.B. (2001). Plant receptor-like kinase gene family: Diversity, function, and signaling. *Science's STKE, 2001*(113), re22.

Shpak, E.D. (2003). Dominant-negative receptor uncovers redundancy in the *Arabidopsis ERECTA* leucine-rich repeat receptor-like kinase signaling pathway that regulates organ shape. *The Plant Cell Online, 15*(5), 1095-1110.

Shukla, D.D., & Ward, C.W. (1988). Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *Journal of General Virology, 69*(11), 2703-2710.

Shulaev, V., Korban, S.S., Sosinski, B., Abbott, A.G., Aldwinckle, H.S., Folta, K.M., Iezzoni, A., Main, D., Arus, P., & Dandekar, A.M. (2008). Multiple models for Rosaceae genomics. *Plant physiology, 147*(3), 985-1003.

Smigocki, A.C., & Hammerschlag, F.A. (1991). Regeneration of plants from peach embryo cells infected with a shooty mutant strain of *Agrobacterium*. *Journal of the American Society for Horticultural Science, 116*(6), 1092-1097.

Socquet-Juglard, D., Kamber, T., Pothier, J.F., Christen, D., Gessler, C., Duffy, B., & Patocchi, A. (2013). Comparative RNA-seq analysis of early-infected peach leaves by the invasive phytopathogen *Xanthomonas arboricola pv. pruni*. *PLoS One, 8*(1), e54196.

Song, W., Wang, G., Chen, L., Kim, H.-S., Pi, L., Holsten, T., Gardner, J., Wang, B., Zhai, W., Zhu, L., Fauquet, C., & Ronald, P. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, by the rice disease resistance gene, *Xa21*. *Science, 270*(5243), 1804.

Song, Z., Guo, S., Zhang, C., Zhang, B., Ma, R., Korir, N.K., & Yu, M. (2015). KT/HAK/KUP potassium transporter genes differentially expressed during fruit development, ripening, and postharvest shelf-life of 'Xiahui6'peaches. *Acta physiologiae plantarum, 37*(7), 131.

Sosinski, B., Gannavarapu, M., Hager, L.D., Beck, L.E., King, G.J., Ryder, C.D., Rajapakse, S., Baird, W.V., Ballard, R.E., & Abbott, A.G. (2000). Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. *Theoretical and Applied Genetics, 101*(3), 421-428.

Spivakov, M., Akhtar, J., Kheradpour, P., Beal, K., Girardot, C., Koscielny, G., Herrero, J., Kellis, M., Furlong, E.E.M., & Birney, E. (2012). Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome biology, 13*(9), R49.

Sumner, A.T., Evans, H.J., & Buckland, R.A. (1973). Mechanisms involved in the banding of chromosomes with quinacrine and Giemsa: I. The effects of fixation in methanol - acetic acid. *Experimental cell research, 81*(1), 214-222.

Sumner, A.T., Taggart, M.H., Mezzanotte, R., & Ferrucci, L. (1990). Patterns of digestion of human chromosomes by restriction endonucleases demonstrated by *in situ* nick translation. *The Histochemical journal, 22*(12), 639-652.

Suzaki, T., Sato, M., Ashikari, M., Miyoshi, M., Nagato, Y., & Hirano, H.-Y. (2004). The gene *FLORAL ORGAN NUMBER1* regulates floral meristem size in rice and encodes a leucine-rich repeat receptor kinase orthologous to *Arabidopsis CLAVATA1*. *Development, 131*(22), 5649.

Symington, L.S. (2002). Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair. *Microbiology and Molecular Biology Reviews, 66*(4), 630.

Szinay, D., Bai, Y., Visser, R., & de Jong, H. (2010). Fish applications for genomics and plant breeding strategies in tomato and other *Solanaceous* crops. *Cytogenetic and Genome Research, 129*(1-3), 199-210.

Tan, B.-C., Joseph, L.M., Deng, W.-T., Liu, L., Li, Q.-B., Cline, K., & McCarty, D.R. (2003). Molecular characterization of the *Arabidopsis* 9-*cis* epoxycarotenoid dioxygenase gene family. *The Plant Journal, 35*(1), 44-56.

Tang, P., Zhang, Y., Sun, X., Tian, D., Yang, S., & Ding, J. (2010). Disease resistance signature of the leucine-rich repeat receptor-like kinase genes in four plant species. *Plant Science, 179*(4), 399-406.

Tatarinova, T.V., Chekalin, E., Nikolsky, Y., Bruskin, S., Chebotarov, D., McNally, K.L., & Alexandrov, N. (2016). Nucleotide diversity analysis highlights functionally important genomic regions. *Sci Rep, 6*, 35730.

Taylor, E.W. (1965). The mechanism of colchicine inhibition of mitosis. *The Journal of Cell Biology, 25*(1), 145.

Thompson, A.J., Jackson, A.C., Symonds, R.C., Mulholland, B.J., Dadswell, A.R., Blake, P.S., Burbidge, A., & Taylor, I.B. (2001). Ectopic expression of a tomato 9-*cis*-epoxycarotenoid dioxygenase gene causes over-production of abscisic acid. *The Plant Journal, 23*(3), 363-374.

Tlaskal, J. (1979). Combined cycloheximide and 8-hydroxyquinoline pretreatment for study of plant chromosomes. *Stain Technology, 54*(6), 313-319.

Tonutti, P., Bonghi, C., Ruperti, B., Tonielli, G.B., & Ramina, A. (1997). Ethylene evolution and 1-aminocyclopropane-1-carboxylate oxidase gene expression during early development and ripening of peach fruit. *Journal of the American Society for Horticultural Science, 122*(5), 642-647.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F., & Komeda, Y. (1996). The Arabidopsis *ERECTA* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell, 8*(4), 735-746.

Torregrosa, L., Fernandez, L., Bouquet, A., Boursiquot, J.M., Pelsy, F., & Martinez-Zapater, J.M. (2011). Origins and consequences of somatic variation in grapevine. *Genetics, genomics, and breeding of grapes*, 68-92.

Toyama, T.K. (1974). Haploidy in peach. *Horticulture Science, 9*, 187-188.

Trainotti, L., Bonghi, C., Ziliotto, F., Zanin, D., Rasori, A., Casadoro, G., Ramina, A., & Tonutti, P. (2006). The use of microarray μPEACH1.0 to investigate transcriptome changes during transition from pre-climacteric to climacteric phase in peach fruit. *Plant Science, 170*(3), 606-613.

Vaistij, F.E., & Jones, L. (2009). Compromised virus-induced gene silencing in RDR6-deficient plants. *Plant physiology, 149*(3), 1399.

Valli, A., Martín-Hernández, A.M., López-Moya, J.J., & García, J.A. (2006). RNA silencing suppression by a second copy of the P1 serine protease of *Cucumber vein yellowing ipomovirus*, a member of the family *Potyviridae* that lacks the cysteine protease HCPro. *Journal of virology, 80*(20), 10055.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., & Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic acids research, 46*(D1), D1190-D1196.

Velculescu, V.E., Zhang, L., Vogelstein, B., & Kinzler, K.W. (1995). Serial analysis of gene expression. *Science, 270*(5235), 484.

Vendramin, E., Pea, G., Dondini, L., Pacheco, I., Dettori, M.T., Gazza, L., Scalabrin, S., Strozzi, F., Tartarini, S., Bassi, D., Verde, I., & Rossini, L. (2014). A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. *PLoS One, 9*(3), e90574.

Verchot, J., Koonin, E.V., & Carrington, J.C. (1991). The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology, 185*(2), 527-535.

Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J., & Cattonaro, F. (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics, 45*(5), 487.

Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C.T., Gasic, K., Micheletti, D., Rosyara, U.R., Cattonaro, F., Vendramin, E., Main, D., Aramini, V., Blas, A.L., Mockler, T.C., Bryant, D.W., Wilhelm, L., Troggio, M., Sosinski, B., Aranzana, M.J., Arús, P., Iezzoni, A., Morgante, M., & Peace, C.P. (2012). Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS One, 7*(4), e35668.

Verde, I., Jenkins, J., Dondini, L., Micali, S., Pagliarani, G., Vendramin, E., Paris, R., Aramini, V., Gazza, L., Rossini, L., Bassi, D., Troggio, M., Shu, S., Grimwood, J., Tartarini, S., Dettori,

M.T., & Schmutz, J. (2017). The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics, 18*(1), 225.

Verde, I., Quarta, R., Cedrola, C., & Dettori, M.T. (2002). *QTL analysis of agronomic traits in a BC1 peach population*.

Walden, R., & Schell, J. (1990). Techniques in plant molecular biology–progress and problems. *The FEBS Journal, 192*(3), 563-576.

Walker, A.R., Lee, E., & Robinson, S.P. (2006). Two new grape cultivars, bud sports of Cabernet Sauvignon bearing pale-coloured berries, are the result of deletion of two regulatory genes of the berry colour locus. *Plant Mol Biol, 62*(4-5), 623-635.

Walker, J.C. (1994). Structure and function of the receptor-like protein kinases of higher plants. *Plant molecular biology, 26*(5), 1599-1609.

Wang, A.M., Doyle, M.V., & Mark, D.F. (1989). Quantitation of mRNA by the polymerase chain reaction. *Proceedings of the National Academy of Sciences, 86*(24), 9717.

Wang, G., Long, Y., Thomma, B.P., de Wit, P.J., Angenent, G.C., & Fiers, M. (2010). Functional analyses of the *CLAVATA2*-like proteins and their domains that contribute to *CLAVATA2* specificity. *Plant Physiol, 152*(1), 320-331.

Wang, L., Zhao, S., Gu, C., Zhou, Y., Zhou, H., Ma, J., Cheng, J., & Han, Y. (2013). Deep RNA-Seq uncovers the peach transcriptome landscape. *Plant molecular biology, 83*(4), 365-377.

Wang, X., Zafian, P., Choudhary, M., & Lawton, M. (1996). The PR5K receptor protein kinase from *Arabidopsis thaliana* is structurally related to a family of plant defense proteins. *Proceedings of the National Academy of Sciences, 93*(6), 2598.

Wang, Y. (1985). Peach growing and germplasm in China. *Acta Horticulturae (Netherlands)*.

Wang, Z., & Zhuang, E. (2001). Chinese Monograph of Fruit Trees. (Peach). In (pp. 326). Beijing: China Forestry Press.

Werner, D.J., & Creller, M.A. (1997). Genetic studies in peach: inheritance of sweet kernel and male sterility. *Journal of the American Society for Horticultural Science, 122*(2), 215-217.

Wills, R.B.H., Scriven, F.M., & Greenfield, H. (1983). Nutrient composition of stone fruit (*Prunus spp*.) cultivars: apricot, cherry, nectarine, peach and plum. *Journal of the Science of Food and Agriculture, 34*(12), 1383-1389.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., & van der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science, 319*(5869), 1527.

Xiao, J., Sekhwal, K.M., Li, P., Ragupathy, R., Cloutier, S., Wang, X., & You, M.F. (2016). Pseudogenes and their genome-wide prediction in plants. *International Journal of Molecular Sciences, 17*(12).

Xie, Y., Ward, R., Fang, C., & Qiao, B. (2007). The urban system in West China: A case study along the mid-section of the ancient Silk Road–He-Xi Corridor. *Cities, 24*(1), 60-73.

Yakushiji, H., Kobayashi, S., Goto-Yamamoto, N., Tae Jeong, S., Sueta, T., Mitani, N., & Azuma, A. (2006). A skin color mutation of grapevine, from black-skinned pinot noir to white-skinned Pinot Blanc, is caused by deletion of the functional *VvmybA1* allele. *Bioscience, Biotechnology, and Biochemistry, 70*(6), 1506-1508.

Yamamoto, T., Shimada, T., Imai, T., Yaegaki, H., Haji, T., Matsuta, N., Yamaguchi, M., & Hayashi, T. (2001). Characterization of morphological traits based on a genetic linkage map in peach. *Breeding Science, 51*(4), 271-278.

Yamamoto, T., Yamaguchi, M., & Hayashi, T. (2005). An Integrated Genetic Linkage Map of Peach by SSR, STS, AFLP and RAPD. *Journal of the Japanese Society for Horticultural Science, 74*(3), 204-213.

Yu, W., Lamb, J.C., Han, F., & Birchler, J.A. (2006). Telomere-mediated chromosomal truncation in maize. *Proceedings of the National Academy of Sciences, 103*(46), 17331.

Zeballos, J.L., Abidi, W., Giménez, R., Monforte, A.J., Moreno, M.Á., & Gogorcena, Y. (2016). Mapping QTLs associated with fruit quality traits in peach [*Prunus persica* (L.) Batsch] using SNP maps. *Tree Genetics & Genomes, 12*(3), 37.

Zhang, C., Bai, M.Y., & Chong, K. (2014). Brassinosteroid-mediated regulation of agronomic traits in rice. *Plant Cell Rep, 33*(5), 683-696.

Zhang, H., & Wing, R.A. (1997). Physical mapping of the rice genome with BACs. In *Oryza: From Molecule to Plant* (pp. 115-127): Springer.

Zhang, H., & Wu, C. (2001). BAC as tools for genome sequencing. *Plant Physiology and Biochemistry, 39*(3-4), 195-209.

Zhang, M., Leng, P., Zhang, G., & Li, X. (2009). Cloning and functional analysis of 9-*cis*-epoxycarotenoid dioxygenase (NCED) genes encoding a key enzyme during abscisic acid biosynthesis from peach and grape fruits. *Journal of Plant Physiology, 166*(12), 1241-1252.

Zhebentyayeva, T.N., Horn, R., Mook, J., Lecouls, A., Georgi, L., Abbott, A.G., Reighard, G.L., Swire-Clark, G., & Baird, W.V. (2006). Physical framework for the peach genome. *Acta horticulturae*.

Zhebentyayeva, T.N., Swire-Clark, G., Georgi, L.L., Garay, L., Jung, S., Forrest, S., Blenda, A.V., Blackmon, B., Mook, J., & Horn, R. (2008). A framework physical map for peach, a model Rosaceae species. *Tree Genetics & Genomes, 4*(4), 745-756.

Zhou, F., Guo, Y., & Qiu, L. (2016). Genome-wide identification and evolutionary analysis of leucine-rich repeat receptor-like protein kinase genes in soybean. *BMC Plant Biol, 16*(1), 58.

Zhou, H., Lin-Wang, K., Wang, H., Gu, C., Dare, A.P., Espley, R.V., He, H., Allan, A.C., & Han, Y. (2015). Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *The Plant Journal, 82*(1), 105-121.

Supplementary material

Supplementary Table 1.S1 Primer pairs used to identify variability in the UDP98-412 region.

Amplicon	Forward primer	Reverse primer	Length	Peach genome v.2 Pp06 Start(v2)	End(v2)	SNPs	Transcript name (peach genome annotation v.2)	IPR description
Amplicon 1	cttgaatctcagtggttgttcg	ttctgaaaggtccacactgg	670	26254140	26254809	–		Leucine-rich repeat
Amplicon 2	ggttccctattgaaaactgtcc	attcaaggatgcaaggtagg	465	26255596	26256060	–	*Prupe.6G281000*	Leucine-rich repeat
Amplicon 3	tgtagattgtgtggtgacagagg	aggagacagaggaaacacaagc	611	26262346	26262956	6		Reverse transcriptase
Amplicon 4	tatgttaaggagcgggttaagg	agtgttccaagttctggtctgg	627	26263412	26264038	6	-	Reverse transcriptase
Amplicon 5	ggattactcaggcaaccatttc	tcccgcaataattgtatccag	646	26270679	26271324	9		Leucine-rich repeat
Amplicon 6	tcccctatcgattgtcaaattc	taatcccacgatggccagaa	551	26271279	26271829	4	*Prupe.6G281100*	Leucine-rich repeat
Amplicon 7	ggggataagttctctttctcagc	ggcctttaatctgattccttcc	473	26276131	26276603	5	*Prupe.6G281200*	Leucine-rich repeat
Amplicon 8	caatttggaaagacctcgaatc	gatagatcaagcacccgaagac	604	26279095	26279698	–		Leucine-rich repeat
Amplicon 9	tccctaacagaggtcaaattcc	gtaacctgggcttttgatatgc	516	26280111	26280626	–	*Prupe.6G281300*	Leucine-rich repeat
UDP98-412(-17K)	atattaccccctcttcgttggt	ctgggtataaaatggggcatct	446	26599970	26600416	–	-	
UDP98-412(-5K)	ggggcatgcacaaacataatag	gcgtcatatatgctgggaagtc	356	26612531	26612887	–	*Prupe.6G287900*	Uncharacterized protein
UDP98-412(-3K)	gtgaataggtttggctctttcc	ccctttcatttacccttgtcc	226	26614309	26614535	–	-	
UDP98-412(-2K)	acttgtaagccgaaagagatgg	agtttacttcacaggccaaagc	422	26614986	26615408	–	-	
UDP98-412(-1.8K)	ttaattccactcctctctcatgc	tccctctcaacataaatgatcc	290	26615542	26615832	–	-	
UDP98-412(-0.2K)	cagcaccactgactaagtgacc	cctaaccgcagctctttatacg	200	26617199	26617399	–	-	
UDP98-412	gccaactgaaaagtctctgtcc	tgccactagatgtgttctgagg	504	26621199	26621703	–	*Prupe.6G288000*	Uncharacterized protein

UDP98-412(+4K)	gccaactgaaaagtctctgtcc	tgccactagatgtgttctgagg	504	26621199	26621703	–	*Prupe.6G288000*	Uncharacterized protein
UDP98-412(+7K)	gagcttacatttcaggagttcg	ctgtaggacacgtttgttttgg	508	26624556	26625064	–		
UDP98-412(+10K)	aatccaggagatgctgtaatgg	ctcttcatcttgtcagctctgg	541	26626955	26627496	–	*Prupe.6G288100*	Uncharacterized protein
UDP98-412(+12K)	aagtccaagtcaaaacgtaggc	gaatgttctccctcatggtagg	587	26629595	26630182	–		
UDP98-412(+15K)	caagaagccaaatcacactgc	ctcatggagggtagatctgagg	677	26632723	26633400	–		
UDP98-412(+20K)	gtcgcaagttgaccatgttacc	atcaaccacgagagtccatagg	680	26636263	26636943	–	-	
UDP98-412(+22K)	atagcttcggtagggtacatgc	tagcctaccccaagaaaatacg	672	26639450	26640122	–	-	
UDP98-412(+25K)	agctgctcaaggagaaagagg	ataactcgtgcgaatctcaagg	569	26642190	26642759	–	-	

Supplementary Table 1.S2 List of samples genotyped with the SSR marker UDP98-412, PC4 and PC3.

Cultivar	Phenotype	UDP98-412 genotype[1]	Small Indels genotype[2] (PC4)	10 Kb indel genotype[3] (PC3)
Almudi[s]	Flat	131/131	464/469	941/1620
Almunia[s]	Flat	127/131	464/470	941/1620
ASF 01-81	Flat	129/131	464/469	941/1620
ASF 02-80	Flat	129/131	464/469	941/1620
ASF 02-83	Flat	129/131	464/469	941/1620
ASF 02-86	Flat	123/131	464/469	941/1620
ASF 02-87	Flat	129/131	464/469	941/1620
ASF 03-81	Flat	129/131	464/469	941/1620
ASF 04-71[s]	Flat	123/131	464/469	941/1620
ASF 04-81[s]	Flat	129/131	464/469	941/1620
ASF 04-92[s]	Flat	129/131	464/469	941/1620
ASF 04-93[s]	Flat	129/131	464/469	941/1620
ASF 04-94[s]	Flat	129/131	464/469	941/1620
ASF 05-81[s]	Flat	129/131	464/469	-
ASF 05-92[s]	Flat	129/131	464/469	941/1620
ASF 05-93[s]	Flat	129/131	464/469	941/1620
ASF 06-71[s]	Flat	127/131	464/469	941/1620
ASF 06-73[s]	Flat	129/131	464/469	941/1620
ASF 06-80[s]	Flat	129/131	464/469	941/1620
ASF 06-83[s]	Flat	131/131	464/469	941/1620
ASF 06-87[s]	Flat	129/131	464/469	941/1620
ASF 06-88[s]	Flat	129/131	464/469	941/1620
ASF 06-90[s]	Flat	127/131	464/469	941/1620
ASF 06-91[s]	Flat	127/131	464/469	941/1620
ASF 06-96[s]	Flat	129/131	464/469	941/1620
ASF 06-97[s]	Flat	129/131	464/469	941/1620
ASF 06-99[s]	Flat	129/131	464/469	941/1620

ASF 07-73[s]	Flat	127/131	464/469	941/1620
ASF 07-78	Flat	129/131	464/469	941/1620
ASF 07-80[s]	Flat	129/131	464/469	941/1620
ASF 07-98[s]	Flat	127/131	464/469	941/1620
Caspe	Flat	131/131	464/469	941/1620
Donutnice[s]	Flat	129/131	464/469	-
EP 93.06	Flat	129/131	464/469	941/1620
Flatelate[s]	Flat	129/131	464/469	941/1620
Green Pan Tao[s]	Flat	129/131	464/469	941/1620
Mesembrine[s]	Flat	123/131	464/469	941/1620
Nebuly	Flat	131/131	-	941/1620
Niqui	Flat	-	464/469	941/1620
Ordigan[s]	Flat	129/131	464/469	941/1620
Oriane[s]	Flat	129/131	464/469	941/1620
Oriola[s]	Flat	123/131	464/469	941/1620
P01F004A003	Flat	129/131	464/469	941/1620
P01F013A049	Flat	123/131	464/469	941/1620
Par_T-Robert[s]	Flat	125/131	464/469	941/1620
Paraguayo Amarillo	Flat	125/131	464/469	941/1620
Paraguayo B	Flat	123/131	464/469	941/1620
Paraguayo Delfin	Flat	131/131	464/469	941/1620
Paraguayo Francia	Flat	125/131	464/469	941/1620
Paraguayo Jota	Flat	127/131	464/469	941/1620
Platibelle	Flat	123/131	464/469	941/1620
Platifur	Flat	-	464/469	941/1620
Platycarpa[s]	Flat	131/131	464/469	941/1620
San Mateo[s]	Flat	131/131	464/469	941/1620
Subirana[s]	Flat	129/131	464/469	941/1620
Sweep Cap[s]	Flat	129/131	464/469	941/1620

UFO 1[s]	Flat	127/131	464/469	941/1620
UFO 2[s]	Flat	127/131	464/469	941/1620
UFO 3[s]	Flat	127/131	464/469	941/1620
UFO 4[s]	Flat	127/131	464/469	941/1620
UFO 5[s]	Flat	123/131	464/469	941/1620
UFO 6[s]	Flat	127/131	464/469	941/1620
UFO 7[s]	Flat	123/131	464/469	941/1620
UFO 8[s]	Flat	129/131	464/469	941/1620
UFO 9[s]	Flat	125/131	464/469	941/1620
Vilamayor[s]	Flat	131/131	464/469	941/1620
Wan Pan Tao[s]	Flat	129/131	464/469	941/1620
Agabés	Round	127/129	469/469	941/941
Alice	Round	125/127	469/469	941/941
ASF 00-01	Round	-	469/469	941/941
ASF 01-03	Round	121/129	469/469	941/941
ASF 01-04	Round	123/123	469/469	941/941
ASF 01-05	Round	123/129	469/469	941/941
ASF 01-29	Round	123/129	469/469	941/941
ASF 02-08	Round	123/131	469/469	941/941
ASF 02-22	Round	127/129	469/469	941/941
ASF 02-23	Round	123/129	469/469	941/941
ASF 02-27	Round	129/129	469/469	941/941
ASF 02-48	Round	129/129	469/469	941/941
ASF 02-52	Round	129/129	469/469	941/941
ASF 02-55	Round	129/129	469/469	941/941
ASF 02-65	Round	129/129	469/469	941/941
ASF 03-02	Round	123/123	469/469	941/941
ASF 03-21	Round	129/129	469/469	941/941
ASF 03-28	Round	129/129	469/469	941/941
ASF 03-62	Round	123/129	469/469	941/941
ASF 03-63	Round	123/123	469/469	941/941

ASF 03-64	Round	123/129	469/469	941/941
ASF 04-06[s]	Round	129/129	469/469	-
ASF 04-09[s]	Round	123/129	469/469	941/941
ASF 04-10	Round	123/129	469/469	941/941
ASF 04-13	Round	123/129	469/469	941/941
ASF 04-14	Round	123/129	469/469	941/941
ASF 04-23[s]	Round	123/129	469/469	941/941
ASF 04-26	Round	127/129	469/469	941/941
ASF 04-27[s]	Round	129/129	469/469	-
ASF 04-30[s]	Round	123/129	469/469	941/941
ASF 04-42[s]	Round	129/129	469/469	941/941
ASF 04-52[s]	Round	123/129	469/469	941/941
ASF 04-53[s]	Round	129/129	469/469	-
ASF 05-01	Round	127/129	469/469	941/941
ASF 05-03	Round	123/127	469/469	941/941
ASF 05-08[s]	Round	127/129	469/469	941/941
ASF 05-15	Round	129/129	469/469	941/941
ASF 05-19	Round	123/123	469/469	941/941
ASF 05-25[s]	Round	127/129	469/469	-
ASF 05-26	Round	-	469/469	941/941
ASF 05-48[s]	Round	129/129	469/469	941/941
ASF 06-07[s]	Round	127/127	469/469	-
ASF 06-19	Round	-	469/469	941/941
ASF 99-02	Round	-	469/469	941/941
Azurite	Round	127/127	469/469	941/941
Bigbel [s]	Round	123/129	469/469	941/941
Bigsun[s]	Round	129/129	469/469	941/941
Bigtop	Round	123/129	469/469	941/941
Binaced [s]	Round	129/129	469/469	941/941
Calabacero[s]	Round	125/127	469/469	941/941
Calante[s]	Round	127/127	469/469	941/941

Catherina[s]	Round	123/127	469/469	941/941
Conserva 458[s]	Round	123/123	469/469	-
Elegant Lady [s]	Round	125/127	469/469	-
EP 94.20	Round	129/129	469/469	941/941
EP 94.28	Round	-	469/469	941/941
EP 97.48	Round	123/129	469/469	941/941
Evaisa[s]	Round	127/127	469/469	941/941
Extreme July[s]	Round	123/129	469/469	941/941
Extreme Red	Round	123/125	469/469	941/941
Extreme Sweet[s]	Round	129/129	469/469	941/941
Feng Bao[s]	Round	129/131	469/469	941/941
Feraude[s]	Round	125/127	469/469	941/941
Fercluse[s]	Round	125/127	469/469	941/941
Ferlot[s]	Round	125/127	469/469	941/941
Garaco	Round	-	469/469	941/941
Garcica[s]	Round	127/129	469/469	941/941
Gardeta	Round	123/127	469/469	941/941
Gartairo	Round	123/123	469/469	941/941
GEM083[s]	Round	129/129	469/469	-
GEM086[s]	Round	129/129	469/469	-
GEM090[s]	Round	129/131	469/469	-
Honey Glo[s]	Round	129/129	469/469	941/941
Honey royale[s]	Round	123/129	469/469	941/941
IFF0331[s]	Round	129/133	469/469	941/941
IFF0800[s]	Round	123/123	469/469	941/941
IFF0813[s]	Round	127/127	469/469	941/941
IFF0962[s]	Round	129/129	469/469	-
IFF1182[s]	Round	127/129	469/469	-
IFF1190[s]	Round	127/129	469/469	941/941
IFF1230[s]	Round	123/129	469/469	941/941
IFF1233[s]	Round	123/129	469/469	-

Indian Freestone[s]	Round	125/125	469/469	-
Jeronimo[s]	Round	120/129	469/469	941/941
Jing Yu[s]	Round	131/131	469/469	941/941
Kawanakajima Hakutou[s]	Round	123/131	469/469	941/941
Kou Ho[s]	Round	129/131	469/469	941/941
Latefair[s]	Round	123/125	469/469	941/941
Luciana[s]	Round	125/129	469/469	941/941
Malatewhite[s]	Round	129/129	469/469	-
Maycrest[s]	Round	120/129	469/469	941/941
Nectabang[s]	Round	123/127	469/469	941/941
Nectajewel[s]	Round	127/129	469/469	-
Nectalady[s]	Round	131/131	469/469	-
Nectapom[s]	Round	129/129	469/469	941/941
Nectariane[s]	Round	129/129	469/469	-
Nectaross [s]	Round	127/127	469/469	-
Nectarreve[s]	Round	127/129	469/469	-
Okayama #11[s]	Round	123/131	469/469	941/941
P01F009A059	Round	123/129	469/469	941/941
P01F011A088	Round	123/129	469/469	941/941
Peregrine[s]	Round	129/131	469/469	941/941
Romea[s]	Round	123/127	469/469	941/941
Stark Redgold	Round	-	-	941/941
Surprise	Round	129/129	469/469	941/941
Villa Giulia[s]	Round	120/129	469/469	941/941
Xin Dai Jiu Bao[s]	Round	125/131	469/469	-
Xiong Yue[s]	Round	131/131	469/469	-
Yumyeong[s]	Round	131/131	469/469	941/941
Zin Dai Jiu Bao[s]	Round	125/131	469/469	941/941
Aborting02[s]	Aborting	131/131	464/464	1620/1620

| Aborting08[s] | Aborting | 131/131 | 464/464 | 1620/1620 |
| Aborting17[s] | Aborting | 131/131 | 464/464 | 1620/1620 |

[1] Allele 131 reported to be associated with the flat phenotype (Picañol *et al*. 2013). Genotypes labeled with in bold represent those scaping for the association.

[2] Primer Combination 4 (PC4) was used to genotype the two small INDELs inside PRUPE.6G284400. Allele 464 is associated with the flat and 469 with the round phenotypes.

[3] Primer Combination 3 (PC3) was used to genotype the 10kb deletion affecting PRUPE.6G284400. Band 941 identifies the round and 1,620 the flat associated alleles.

[s] Varieties sequenced to validate Amplicon5 and Amplicon6 polymorphisms.

Supplementary Table 1.S3 Significant alignments found through the BLAST analysis of *Prupe.6G281100* translated protein against the PLAZA protein sequence database.

Sequences producing significant alignments	Homologous	Orthologous	Bits	Value	Sp	Description (annoMine)
PPE_006G28800	HOM03D000009	ORTHO03D000261	1498	0	*P persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_007G11060	HOM03D000009	ORTHO03D000261	812	0	*P persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_006G28780	HOM03D000009	ORTHO03D035040	708	0	*P persica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD04G012840	HOM03D000009	ORTHO03D000539	677	0	*Malus x domiestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
PPE_008G06340	HOM03D000009	ORTHO03D000261	665	0	*P persica*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G28810	HOM03D000009	ORTHO03D018352	662	0	*P persica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G077620	HOM03D000009	ORTHO03D000539	642	0	*Malus x domiestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G095710	HOM03D000009	ORTHO03D000539	638	0	*Malus x domiestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G452650	HOM03D000009	ORTHO03D000539	630	0	*Malus x domiestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G088240	HOM03D000009	ORTHO03D023912	610	0	*Malus x domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
FV6G27840	HOM03D000009	ORTHO03D092018	609	0	*Fragaria Vesca*	LRR receptor-like serine/threonine-protein kinase GSO2, Precursor (probable)/receptor-like kinase protein THICK TASSEL DWARF1
MD00G235490	HOM03D000009	ORTHO03D057804	601	0	*Malus x domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
FV6G10690	HOM03D000009	ORTHO03D018352	593	0	*Fragaria Vesca*	LRR receptor-like serine/threonine-protein kinase GSO1, Precursor (probable)/Description (AnnoMine) receptor-like kinase protein THICK TASSEL DWARF1

FV0G24790	HOM03D000009	ORTHO03D018352	588	0	*Fragaria Vesca*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G095690	HOM03D000009	ORTHO03D000539	587	0	*Malus x domiestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
FV7G08990	HOM03D000009	ORTHO03D000261	585	0		LRR receptor-like serine/threonine-protein kinase GSO1, Precursor/Description (AnnoMine) receptor-like kinase protein FLORAL ORGAN NUMBER1
PPE_006G29520	HOM03D000009	ORTHO03D028407	584	0	*P persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G282960	HOM03D000009	ORTHO03D000539	583	0	*Malus x domiestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD04G012900	HOM03D000009	ORTHO03D000261	578	0	*Malus x domiestica*	receptor-like kinase protein THICK TASSEL DWARF1
FV6G10050	HOM03D000009	ORTHO03D018352	576	0	*Fragaria Vesca*	LRR receptor-like serine/threonine-protein kinase GSO1, Precursor (probable)/Description (AnnoMine) receptor-like kinase protein THICK TASSEL DWARF1
MD00G095720	HOM03D000009	ORTHO03D000539	569	0	*Malus x domiestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD12G018180	HOM03D000009	ORTHO03D000539	561	0	*Malus x domiestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD08G006770	HOM03D000009	ORTHO03D028407	550	5.00E-179		No description found
FV6G09780	HOM03D000009	ORTHO03D018352	553	6.00E-179	*Fragaria Vesca*	LRR receptor-like serine/threonine-protein kinase GSO1, Precursor (probable)/Description (AnnoMine) receptor-like kinase protein THICK TASSEL DWARF1
MD12G018290	HOM03D000009	ORTHO03D035040	546	3.00E-175		Brassinosteroid LRR receptor kinase
FV0G24860	HOM03D000009	ORTHO03D018352	533	2.00E-173	*Fragaria Vesca*	LRR receptor-like serine/threonine-protein kinase GSO1, Precursor (probable)/Description (AnnoMine) receptor-like kinase protein THICK TASSEL DWARF1

MD12G018370	HOM03D000009	ORTHO03D000539	535	3.00E-172		receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G109640	HOM03D000009	ORTHO03D000539	524	4.00E-170		receptor-like kinase protein THICK TASSEL DWARF1
MD00G282930	HOM03D000009	ORTHO03D000539	518	1.00E-168		inactive leucine-rich repeat receptor-like protein kinase At2g25790
MD00G088230	HOM03D000009	ORTHO03D000261	516	4.00E-167		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G235410	HOM03D000009	ORTHO03D089453	486	3.00E-164		No description found
MT4G017730	HOM03D000009	ORTHO03D000539	506	4.00E-164		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PT16G12050	HOM03D000009	ORTHO03D001987	502	5.00E-162		receptor-like kinase protein THICK TASSEL DWARF1
FV4G16710	HOM03D000009	ORTHO03D018352	497	5.00E-160	*Fragaria Vesca*	LRR receptor-like serine/threonine-protein kinase GSO1, Precursor (probable)/Description (AnnoMine) receptor-like kinase protein THICK TASSEL DWARF1
TC0005G01150	HOM03D000009	ORTHO03D001987	497	2.00E-159		No description found
GM03G07400	HOM03D000009	ORTHO03D000261	486	4.00E-158		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD08G006780	HOM03D000009	ORTHO03D000539	488	6.00E-157		receptor-like kinase protein FLORAL ORGAN NUMBER1
FV6G10040	HOM03D000009	ORTHO03D018352	489	1.00E-155	*Fragaria Vesca*	LRR receptor-like serine/threonine-protein kinase GSO1, Precursor (probable)/Description (AnnoMine) receptor-like kinase protein THICK TASSEL DWARF1
GR12G11700	HOM03D000009	ORTHO03D001987	486	1.00E-155		receptor-like kinase protein THICK TASSEL DWARF1
MD00G351490	HOM03D000009	ORTHO03D001987	478	4.00E-154		receptor-like kinase protein THICK TASSEL DWARF1
MT4G017720	HOM03D000009	ORTHO03D000539	481	3.00E-153		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G331830	HOM03D000009	ORTHO03D098015	472	4.00E-153		No description found

CS01117G00010	HOM03D000009	ORTHO03D000539	478	1.00E-152	receptor-like kinase protein THICK TASSEL DWARF1
CS00098G00490	HOM03D000009	ORTHO03D001987	477	5.00E-152	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G198270	HOM03D000009	ORTHO03D000261	471	2.00E-151	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
CS00098G00500	HOM03D000009	ORTHO03D001987	476	7.00E-151	receptor-like kinase protein THICK TASSEL DWARF1
ME02042G00020	HOM03D000009	ORTHO03D001987	470	2.00E-149	inactive leucine-rich repeat receptor-like protein kinase At3g28040
CS01062G00010	HOM03D000009	ORTHO03D000539	468	3.00E-149	receptor-like kinase protein THICK TASSEL DWARF1
CS00098G00470	HOM03D000009	ORTHO03D001987	469	3.00E-149	receptor-like kinase protein THICK TASSEL DWARF1
GR02G25940	HOM03D000009	ORTHO03D001987	469	1.00E-148	receptor-like kinase protein THICK TASSEL DWARF1
GR02G26450	HOM03D000009	ORTHO03D001987	468	2.00E-148	receptor-like kinase protein FLORAL ORGAN NUMBER1
GM01G31711	HOM03D000009	ORTHO03D000539	469	2.00E-148	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ME02042G00010	HOM03D000009	ORTHO03D001987	468	2.00E-148	receptor-like kinase protein FLORAL ORGAN NUMBER1
TC0005G01160	HOM03D000009	ORTHO03D001987	465	3.00E-147	No description found
FV6G27830	HOM03D000009	ORTHO03D000261	457	6.00E-147	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
RC29601G00090	HOM03D000009	ORTHO03D001987	460	2.00E-145	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR02G26230	HOM03D000009	ORTHO03D001987	461	2.00E-145	receptor-like kinase protein THICK TASSEL DWARF1
CS01786G00010	HOM03D000009	ORTHO03D001987	459	3.00E-145	receptor-like kinase protein THICK TASSEL DWARF1
MT2G078260	HOM03D000009	ORTHO03D000261	456	5.00E-145	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD04G012850	HOM03D000009	ORTHO03D000539	459	2.00E-144	receptor-like kinase protein THICK TASSEL DWARF1
PT16G12060	HOM03D000009	ORTHO03D001987	457	9.00E-144	receptor-like kinase protein THICK TASSEL DWARF1

GR02G25950	HOM03D000009	ORTHO03D001987	453	4.00E-143	receptor-like kinase protein THICK TASSEL DWARF1
GM03G07240	HOM03D000009	ORTHO03D000539	454	7.00E-143	receptor-like kinase protein THICK TASSEL DWARF1
GR06G18410	HOM03D000009	ORTHO03D001987	454	1.00E-142	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017640	HOM03D000009	ORTHO03D000539	454	1.00E-142	receptor-like kinase protein THICK TASSEL DWARF1
GR12G11720	HOM03D000009	ORTHO03D001987	452	3.00E-142	Brassinosteroid LRR receptor kinase
MD00G485210	HOM03D000009	ORTHO03D000261	428	9.00E-141	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
CP00120G00230	HOM03D000009	ORTHO03D001987	449	2.00E-140	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
CP00120G00220	HOM03D000009	ORTHO03D001987	446	5.00E-140	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04520	HOM03D000009	ORTHO03D000539	442	7.00E-140	receptor-like kinase protein THICK TASSEL DWARF1
GR02G26460	HOM03D000009	ORTHO03D000261	434	9.00E-140	receptor-like kinase protein THICK TASSEL DWARF1
CM00011G00590	HOM03D000009	ORTHO03D000539	444	1.00E-139	receptor-like kinase protein THICK TASSEL DWARF1
MD00G331840	HOM03D000009	ORTHO03D052204	425	2.00E-139	No description found
PPE_006G28820	HOM03D000009	ORTHO03D339783	426	2.00E-139	No description found
EG0001G04420	HOM03D000009	ORTHO03D000539	442	4.00E-139	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G282920	HOM03D000009	ORTHO03D000261	439	6.00E-139	receptor-like kinase protein THICK TASSEL DWARF1
GM07G08770	HOM03D000009	ORTHO03D000539	443	1.00E-138	receptor-like kinase protein FLORAL ORGAN NUMBER1
GM03G18170	HOM03D000009	ORTHO03D000539	441	2.00E-138	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G28840	HOM03D000009	ORTHO03D000261	429	3.00E-138	receptor-like kinase protein THICK TASSEL DWARF1
CL10G20240	HOM03D000009	ORTHO03D000539	442	3.00E-138	receptor-like kinase protein THICK TASSEL DWARF1
GM18G43630	HOM03D000009	ORTHO03D000539	440	2.00E-137	receptor-like kinase protein THICK TASSEL DWARF1

ME04115G00010	HOM03D000009	ORTHO03D001987	439	1.00E-136	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017700	HOM03D000009	ORTHO03D000539	434	1.00E-135	inactive leucine-rich repeat receptor-like protein kinase At3g28040
MT4G013315	HOM03D000009	ORTHO03D000539	429	7.00E-134	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04530	HOM03D000009	ORTHO03D000539	430	7.00E-134	inactive leucine-rich repeat receptor-like protein kinase At2g25790
GM18G43490	HOM03D000009	ORTHO03D000261	425	9.00E-134	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04380	HOM03D000009	ORTHO03D000539	429	1.00E-133	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_008G06330	HOM03D000009	ORTHO03D000261	422	4.00E-133	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017710	HOM03D000009	ORTHO03D000539	424	9.00E-133	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04500	HOM03D000009	ORTHO03D000539	425	2.00E-132	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GM18G43500	HOM03D000009	ORTHO03D000539	424	3.00E-132	receptor-like kinase protein THICK TASSEL DWARF1
LJ2G010310	HOM03D000009	ORTHO03D000261	429	5.00E-132	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G04560	HOM03D000009	ORTHO03D000539	426	9.00E-132	receptor-like kinase protein THICK TASSEL DWARF1
ME04305G00130	HOM03D000009	ORTHO03D000261	420	3.00E-131	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04360	HOM03D000009	ORTHO03D000539	418	1.00E-129	receptor-like kinase protein THICK TASSEL DWARF1
GM18G43510	HOM03D000009	ORTHO03D000539	420	1.00E-129	receptor-like kinase protein THICK TASSEL DWARF1
MD00G279640	HOM03D000009	ORTHO03D000261	413	3.00E-129	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017780	HOM03D000009	ORTHO03D000261	411	1.00E-128	disease resistance family protein/LRR protein
GM07G18640	HOM03D000009	ORTHO03D000539	409	2.00E-127	receptor-like kinase protein THICK TASSEL DWARF1

GM01G29570	HOM03D000009	ORTHO03D000539	414	2.00E-127	receptor-like kinase protein THICK TASSEL DWARF1
GM18G43520	HOM03D000009	ORTHO03D000539	412	2.00E-127	receptor-like kinase protein THICK TASSEL DWARF1
LJ1G031850	HOM03D000009	ORTHO03D000539	411	1.00E-126	receptor-like kinase protein FLORAL ORGAN NUMBER1
SL01G098690	HOM03D000009	ORTHO03D000261	406	3.00E-126	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
CL10G20230	HOM03D000009	ORTHO03D000539	409	3.00E-126	receptor-like kinase protein THICK TASSEL DWARF1
MT4G018940	HOM03D000009	ORTHO03D000539	408	5.00E-126	receptor-like kinase protein FLORAL ORGAN NUMBER1
CM00011G00580	HOM03D000009	ORTHO03D000539	407	3.00E-125	receptor-like kinase protein THICK TASSEL DWARF1
ST01G036370	HOM03D000009	ORTHO03D000261	400	6.00E-124	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0001G04550	HOM03D000009	ORTHO03D028407	401	2.00E-123	No description found
GR02G26440	HOM03D000009	ORTHO03D000261	395	4.00E-123	receptor-like kinase protein THICK TASSEL DWARF1
GM07G18590	HOM03D000009	ORTHO03D000261	396	6.00E-123	receptor-like kinase protein THICK TASSEL DWARF1
ST01G036380	HOM03D000009	ORTHO03D001987	397	7.00E-123	receptor-like kinase protein THICK TASSEL DWARF1
CS01062G00050	HOM03D000009	ORTHO03D000261	387	5.00E-122	receptor-like kinase protein THICK TASSEL DWARF1
ST09G012330	HOM03D000009	ORTHO03D000539	399	1.00E-121	receptor-like kinase protein THICK TASSEL DWARF1
ST09G012370	HOM03D000009	ORTHO03D000261	385	4.00E-121	Brassinosteroid LRR receptor kinase
GM01G29030	HOM03D000009	ORTHO03D000539	395	1.00E-120	receptor-like kinase protein THICK TASSEL DWARF1
VV08G06910	HOM03D000009	ORTHO03D088883	405	2.00E-120	No description found
MT4G417260	HOM03D000009	ORTHO03D000539	395	2.00E-120	receptor-like kinase protein FLORAL ORGAN NUMBER1
GM18G43621	HOM03D000009	ORTHO03D000539	392	1.00E-119	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04340	HOM03D000009	ORTHO03D000539	384	7.00E-119	Brassinosteroid LRR receptor kinase

GM01G28936	HOM03D000009	ORTHO03D000539	389	2.00E-118	receptor-like kinase protein THICK TASSEL DWARF1
MT4G019030	HOM03D000009	ORTHO03D000539	386	4.00E-118	receptor-like kinase protein THICK TASSEL DWARF1
MT4G019010	HOM03D000009	ORTHO03D000539	385	1.00E-117	receptor-like kinase protein THICK TASSEL DWARF1
GM01G29615	HOM03D000009	ORTHO03D000539	389	1.00E-117	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017600	HOM03D000009	ORTHO03D000539	387	2.00E-117	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G28830	HOM03D000009	ORTHO03D028407	382	5.00E-117	No description found
EG0001G04430	HOM03D000009	ORTHO03D000539	383	5.00E-117	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04390	HOM03D000009	ORTHO03D000539	384	7.00E-117	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ST09G012340	HOM03D000009	ORTHO03D000539	384	9.00E-117	receptor-like kinase protein THICK TASSEL DWARF1
SL09G005090	HOM03D000009	ORTHO03D000539	383	2.00E-116	receptor-like kinase protein THICK TASSEL DWARF1
SL09G005080	HOM03D000009	ORTHO03D000539	384	5.00E-116	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017490	HOM03D000009	ORTHO03D000261	375	2.00E-115	receptor-like kinase protein THICK TASSEL DWARF1
BV5G09940	HOM03D000009	ORTHO03D000539	380	6.00E-115	Receptor-like protein kinase BRI1-like 3
LJ2G010060	HOM03D000009	ORTHO03D000539	377	2.00E-114	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04450	HOM03D000009	ORTHO03D000539	373	1.00E-113	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G05420	HOM03D000009	ORTHO03D000539	373	2.00E-113	receptor-like kinase protein THICK TASSEL DWARF1
MD05G014070	HOM03D000009	ORTHO03D000539	372	2.00E-113	receptor-like kinase protein THICK TASSEL DWARF1
GM01G29580	HOM03D000009	ORTHO03D000539	375	4.00E-113	receptor-like kinase protein THICK TASSEL DWARF1
TC0007G10300	HOM03D000009	ORTHO03D014855	368	4.00E-113	No description found
SL10G076500	HOM03D000009	ORTHO03D000261	367	2.00E-112	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

SL01G098680	HOM03D000009	ORTHO03D000261	368	5.00E-112		receptor-like kinase protein THICK TASSEL DWARF1
ST08G023730	HOM03D000009	ORTHO03D020554	364	7.00E-112		No description found
PT00G14700	HOM03D000009	ORTHO03D002896	358	7.00E-110	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PT12G02800	HOM03D000009	ORTHO03D002896	358	9.00E-110	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0001G05510	HOM03D000009	ORTHO03D000539	359	5.00E-108		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ST10G018890	HOM03D000009	ORTHO03D023912	358	8.00E-108		No description found
GR02G26240	HOM03D000009	ORTHO03D000261	353	8.00E-108		receptor-like kinase protein FLORAL ORGAN NUMBER1
CS00019G01260	HOM03D000009	ORTHO03D001987	357	6.00E-107		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0001G05460	HOM03D000009	ORTHO03D000539	354	3.00E-106		receptor-like kinase protein THICK TASSEL DWARF1
ST12G031970	HOM03D000009	ORTHO03D081567	348	2.00E-105		No description found
ST01G006170	HOM03D000009	ORTHO03D018926	346	5.00E-105		No description found
VV12G08270	HOM03D000009	ORTHO03D014855	346	9.00E-105		No description found
GM14G04710	HOM03D000009	ORTHO03D090296	346	1.00E-104		No description found
EG0001G05440	HOM03D000009	ORTHO03D000539	346	1.00E-103		receptor-like kinase protein THICK TASSEL DWARF1
PT12G00940	HOM03D000009	ORTHO03D002896	343	2.00E-103	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
VV19G08190	HOM03D000009	ORTHO03D018926	340	2.00E-103		No description found
SL01G005780	HOM03D000009	ORTHO03D014465	341	2.00E-103		No description found
CL02G02830	HOM03D000009	ORTHO03D052129	341	4.00E-103		No description found
PT12G02530	HOM03D000009	ORTHO03D002896	342	6.00E-103	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

PT12G02900	HOM03D000009	ORTHO03D002896	340	1.00E-102	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ST10G018880	HOM03D000009	ORTHO03D000261	341	1.00E-102		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR02G06060	HOM03D000009	ORTHO03D091861	337	9.00E-102		No description found
EG0006G05740	HOM03D000009	ORTHO03D018926	336	3.00E-101		No description found
PT12G02760	HOM03D000009	ORTHO03D082383	340	5.00E-101		No description found
GM14G34930	HOM03D000009	ORTHO03D048478	333	1.00E-100		No description found
MD09G008030	HOM03D000009	ORTHO03D058940	325	2.00E-100		No description found
TC0026G00060	HOM03D000009	ORTHO03D014855	338	4.00E-100		No description found
ST10G018870	HOM03D000009	ORTHO03D000261	332	6.00E-100		receptor-like kinase protein THICK TASSEL DWARF1
OS01G06760	HOM03D000009	ORTHO03D052129	331	6.00E-100		No description found
OS12G11860	HOM03D000009	ORTHO03D000261	337	7.00E-100		receptor-like kinase protein FLORAL ORGAN NUMBER1
BR07G23090	HOM03D000009	ORTHO03D013283	330	8.00E-100		No description found
ATR_00029G03900	HOM03D000009	ORTHO03D000539	338	9.00E-100		receptor-like kinase protein FLORAL ORGAN NUMBER1
PT00G11830	HOM03D000009	ORTHO03D091433	333	1.00E-99		No description found
ST10G018600	HOM03D000009	ORTHO03D023912	335	2.00E-99		No description found
ME08617G00010	HOM03D000009	ORTHO03D000261	335	2.00E-99		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
CRU_003G22220	HOM03D000009	ORTHO03D001279	330	3.00E-99		No description found
PT12G01060	HOM03D000009	ORTHO03D002896	332	3.00E-99	*Populus tricocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
MT5G087320	HOM03D000009	ORTHO03D014880	335	7.00E-99		No description found

PT01G43770	HOM03D000009	ORTHO03D001987	334	9.00E-99		receptor-like kinase protein THICK TASSEL DWARF1
TC0007G11860	HOM03D000009	ORTHO03D020746	334	2.00E-98		No description found
EG0002G05890	HOM03D000009	ORTHO03D054184	326	3.00E-98		No description found
OS12G12120	HOM03D000009	ORTHO03D000261	333	3.00E-98		receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G12010	HOM03D000009	ORTHO03D000261	332	4.00E-98		receptor-like kinase protein FLORAL ORGAN NUMBER1
PT12G00560	HOM03D000009	ORTHO03D002896	327	4.00E-98	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS12G11370	HOM03D000009	ORTHO03D000261	332	6.00E-98		receptor-like kinase protein THICK TASSEL DWARF1
OS12G11720	HOM03D000009	ORTHO03D000261	332	6.00E-98		receptor-like kinase protein FLORAL ORGAN NUMBER1
OS01G04070	HOM03D000009	ORTHO03D000261	335	6.00E-98		receptor-like kinase protein FLORAL ORGAN NUMBER1
ATR_00029G03770	HOM03D000009	ORTHO03D000539	333	9.00E-98		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ME08617G00030	HOM03D000009	ORTHO03D100620	332	9.00E-98		No description found
ST04G005580	HOM03D000009	ORTHO03D000980	328	9.00E-98		No description found
AT3G05660	HOM03D000009	ORTHO03D019070	329	9.00E-98		No description found
OS01G06730	HOM03D000009	ORTHO03D000261	332	1.00E-97		receptor-like kinase protein FLORAL ORGAN NUMBER1
PT01G12840	HOM03D000009	ORTHO03D001987	331	2.00E-97		receptor-like kinase protein THICK TASSEL DWARF1
CS00735G00010	HOM03D000009	ORTHO03D014855	321	3.00E-97		No description found
BR08G11670	HOM03D000009	ORTHO03D013283	323	3.00E-97		No description found
MT5G086945	HOM03D000009	ORTHO03D023777	323	4.00E-97		No description found
OS12G12130	HOM03D000009	ORTHO03D000261	329	7.00E-97		receptor-like kinase protein FLORAL ORGAN NUMBER1

TC0004G03660	HOM03D000009	ORTHO03D101367	322	8.00E-97		No description found
TC0007G11150	HOM03D000009	ORTHO03D028397	322	1.00E-96		No description found
BR04G14920	HOM03D000009	ORTHO03D001279	325	1.00E-96		No description found
PT12G00920	HOM03D000009	ORTHO03D002896	325	2.00E-96	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PT11G14110	HOM03D000009	ORTHO03D018915	319	2.00E-96		No description found
EG0007G24850	HOM03D000009	ORTHO03D023268	319	2.00E-96		No description found
BR01G38030	HOM03D000009	ORTHO03D001279	324	4.00E-96		No description found
AT3G28890	HOM03D000009	ORTHO03D013283	320	7.00E-96		No description found
ST12G008210	HOM03D000009	ORTHO03D035944	317	1.00E-95		No description found
TC0007G14890	HOM03D000009	ORTHO03D049987	322	3.00E-95		No description found
GR01G22340	HOM03D000009	ORTHO03D000261	323	4.00E-95		receptor-like kinase protein FLORAL ORGAN NUMBER1
GR02G05290	HOM03D000009	ORTHO03D000980	322	6.00E-95		No description found
CRU_003G22330	HOM03D000009	ORTHO03D001279	318	8.00E-95		No description found
OS12G10870	HOM03D000009	ORTHO03D000261	323	8.00E-95		receptor-like kinase protein FLORAL ORGAN NUMBER1
PT01G38910	HOM03D000009	ORTHO03D094790	322	1.00E-94		No description found
BV0G79990	HOM03D000009	ORTHO03D000261	323	1.00E-94		receptor-like kinase protein FLORAL ORGAN NUMBER1
ST12G005940	HOM03D000009	ORTHO03D000980	320	1.00E-94		No description found
SL04G054450	HOM03D000009	ORTHO03D017180	320	2.00E-94		No description found

EG0006G04500	HOM03D000009	ORTHO03D091457	317	2.00E-94		No description found
PT11G05500	HOM03D000009	ORTHO03D002896	316	2.00E-94	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS12G11930	HOM03D000009	ORTHO03D000261	322	3.00E-94		receptor-like kinase protein FLORAL ORGAN NUMBER1
BR04G19410	HOM03D000009	ORTHO03D008455	316	3.00E-94		No description found
MT4G018910	HOM03D000009	ORTHO03D000261	321	4.00E-94		receptor-like kinase protein THICK TASSEL DWARF1
SL01G005730	HOM03D000009	ORTHO03D014465	317	1.00E-93		No description found
GR09G35460	HOM03D000009	ORTHO03D012248	320	1.00E-93		No description found
ST05G014920	HOM03D000009	ORTHO03D020554	314	1.00E-93		No description found
RC29836G00110	HOM03D000009	ORTHO03D000980	319	2.00E-93		No description found
ST12G031890	HOM03D000009	ORTHO03D020553	312	3.00E-93		No description found
PT12G02740	HOM03D000009	ORTHO03D002896	318	3.00E-93	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G351530	HOM03D000009	ORTHO03D000539	316	4.00E-93		receptor-like kinase protein THICK TASSEL DWARF1
OS12G11680	HOM03D000009	ORTHO03D000261	319	6.00E-93		receptor-like kinase protein FLORAL ORGAN NUMBER1
CS00034G00160	HOM03D000009	ORTHO03D000980	314	7.00E-93		No description found
TC0007G00430	HOM03D000009	ORTHO03D020746	314	9.00E-93		No description found
PT01G43760	HOM03D000009	ORTHO03D018915	309	1.00E-92		No description found
OS01G06520	HOM03D000009	ORTHO03D000261	318	1.00E-92		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0006G04610	HOM03D000009	ORTHO03D081793	313	1.00E-92		No description found

ST08G023130	HOM03D000009	ORTHO03D086207	313	2.00E-92		No description found
CS01401G00010	HOM03D000009	ORTHO03D035292	316	2.00E-92		No description found
CRU_004G03270	HOM03D000009	ORTHO03D013283	310	4.00E-92		No description found
TC0007G00380	HOM03D000009	ORTHO03D020746	309	5.00E-92		No description found
SL12G099870	HOM03D000009	ORTHO03D084524	309	5.00E-92		No description found
MD00G279620	HOM03D000009	ORTHO03D035040	295	6.00E-92		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
VV18G14740	HOM03D000009	ORTHO03D018926	306	7.00E-92		No description found
RC29601G00080	HOM03D000009	ORTHO03D000539	316	7.00E-92		receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0513G00020	HOM03D000009	ORTHO03D004402	311	8.00E-92		No description found
OS12G11500	HOM03D000009	ORTHO03D000261	315	8.00E-92		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
BR07G02580	HOM03D000009	ORTHO03D042977	309	1.00E-91		No description found
PT12G02690	HOM03D000009	ORTHO03D002896	310	1.00E-91	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0013G00140	HOM03D000009	ORTHO03D004402	310	1.00E-91		No description found
PT00G14580	HOM03D000009	ORTHO03D084117	316	1.00E-91		No description found
MT5G087090	HOM03D000009	ORTHO03D014880	315	2.00E-91		No description found
OS04G40440	HOM03D000009	ORTHO03D000261	315	2.00E-91		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
BR04G19400	HOM03D000009	ORTHO03D008455	309	3.00E-91		No description found
PT12G02540	HOM03D000009	ORTHO03D002896	313	3.00E-91	*Populus tricocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

Supplementary Table 1.S4 Primer pairs used for the whole sequencing *Prupe.6G281100* (PC1 and PC2) and for genotyping (PC3 and PC4) round and flat cultivars.

Primer combination	Forward primer name	Short name	Sequence (5'-3')	Reverse primer name	Short name	Sequence (5'-3)	Start	End	Fragment size* (bp)
PC1	Prupe.6G281100_5PrimF	Prupe11_F	ATTTCTTGCAGGCACCGACT	Prupe.6G281100_3PrimR	Prupe11_R	AGTCCATCTGTCGAGTTGGC	26269254	26272587	3,333
PC2	Prupe.6G281100(-10K)_F	IndelS_F	TCCACCACGCCTTATCTGAC				26259705		12882
	Prupe.6G281100_1F		CCACCACAACCTTTATTTCTC	Prupe.6G281100_1R		GAGACTGCTTGAATCGTCAATG	26269692	26270082	390
	Prupe.6G281100_2F		GCCTTCAATTTTCTCATGATCC	Prupe.6G281100_2R		ATCTGGTTTTCTGAAAGGTCCA	26269904	26270498	594
	Prupe.6G281100_3F		TGACAACCTACTTGAGGGGAGT	Prupe.6G281100_3R		ACCACCTAACTGATTTCCATCG	26270343	26271000	657
	Prupe.6G281100_4F		GGATTACTCAGGCAACCATTTC	Prupe.6G281100_4R	IndelS_R	TCCCGCAATAATTGTATCCAG	26270679	26271324	645
	Prupe.6G281100_5F		TCCTTGTGTTGCGGTCCAACA	Prupe.6G281100_5R		TAATCCCACGATGGCCAGAA	26271113	26271829	716
	Prupe.6G281100_6F		CACCTTGATTGACTTCTCTTGC	Prupe.6G281100_6R		TAAAGAAAAAGATGGCCAGGAA	26271417	26272102	685

Primer combination	Forward primer name	Short name	Sequence (5'-3')	Reverse primer name	Short name	Sequence (5'-3)	Start	End	PCR band Size (bp)
PC3	Prupe.6G281100(-10K)_F	IndelS_F	TCCACCACGCCTTATCTGAC	Prupe.6G281100_4R	IndelS_R	AGTCCATCTGTCGAGTTGGC			1620 (flat)
				IndelS_2R		TTTCCAAGCTTTAGGACGAGGA			941 (round)
PC4	FlatIn_F		ATTATTCCCCCATGCTTGAC	Prupe.6G281100_4R	IndelS_R				464 (flat)/469 (round)

*Expected size (in bp) based on the peach reference genome v2.0 (Verde *et al*. 2017)

Supplementary Table 2.S1 Primers used for PCR and sequencing.

Primer name	Position of 5' start (GDR v2.0.a1)	Sequence (5' -> 3')	Tm (ºC)
Prupe14_1F	26278900	GTGGCATTGACAATTCCAGC	58
Prupe14_1R	26288029	AGAGGGAGTGAAGCTTCTGT	58
Prupe14_2F	26280881	GCCTGACCATGCAATTTCAC	58
Prupe14_2R	26290754	TCTGCTGTGAGTCTGTGAGA	58
Prupe14_3F	26279495	CCAACTCCAGGGAAAACTCC	58
Prupe14_4F	26280008	ACGCATTGAATGTCCCAAGA	58
Prupe14_5F	26280604	AGCATATCAAAAGCCCAGGT	57
Prupe14_3R	26287475	GCAGAGTCAAATTGGAGGCT	58
Prupe14_NDF	26280832	AGTTCACGGCTGGTTTTCTT	58
Prupe14_NR	26287312	GCTGGCTGTGATATGTGACA	58
Prupe14_NintF	26287118	CAGTCTAACTTTGTCTACGC	55
Prupe14_aR	26290304	CAACACAAGTGAGCCAATGG	57
Prupe14_bR	26289842	GCAAGAGAGGTCAATCGAGG	58
Prupe14_cR	26289332	AGAACCTGAAGATTTGACGCA	58
Prupe14_dR	26288781	GCCGCTGAAGTTGTTTGAAG	58
Prupe14_eR	26288315	AACATTTTGAGGTCGCCCAT	58
Prupe14_RTF	26287853	TTGGCAACCTTTCTGAGCTT	58
Prupe14_RTR	26288029	AGAGGGAGTGAAGCTTCTGT	58
Prupe14_sRTF	26288376	CCTTACACAATTGGTTTCTGTG	57
Prupe14_sRTR	26288510	TGAGAGGAATTAATCTGACCTG	57
Prupe15_F	26292876	AAACGCGGCTCTGAGTCTTC	56
Prupe15_R	26296643	TGTCCCACCGTTTTGGGTTG	56
Prupe14_dwR	26292817	AGCTTTCTGACAGTCACTCCTG	57
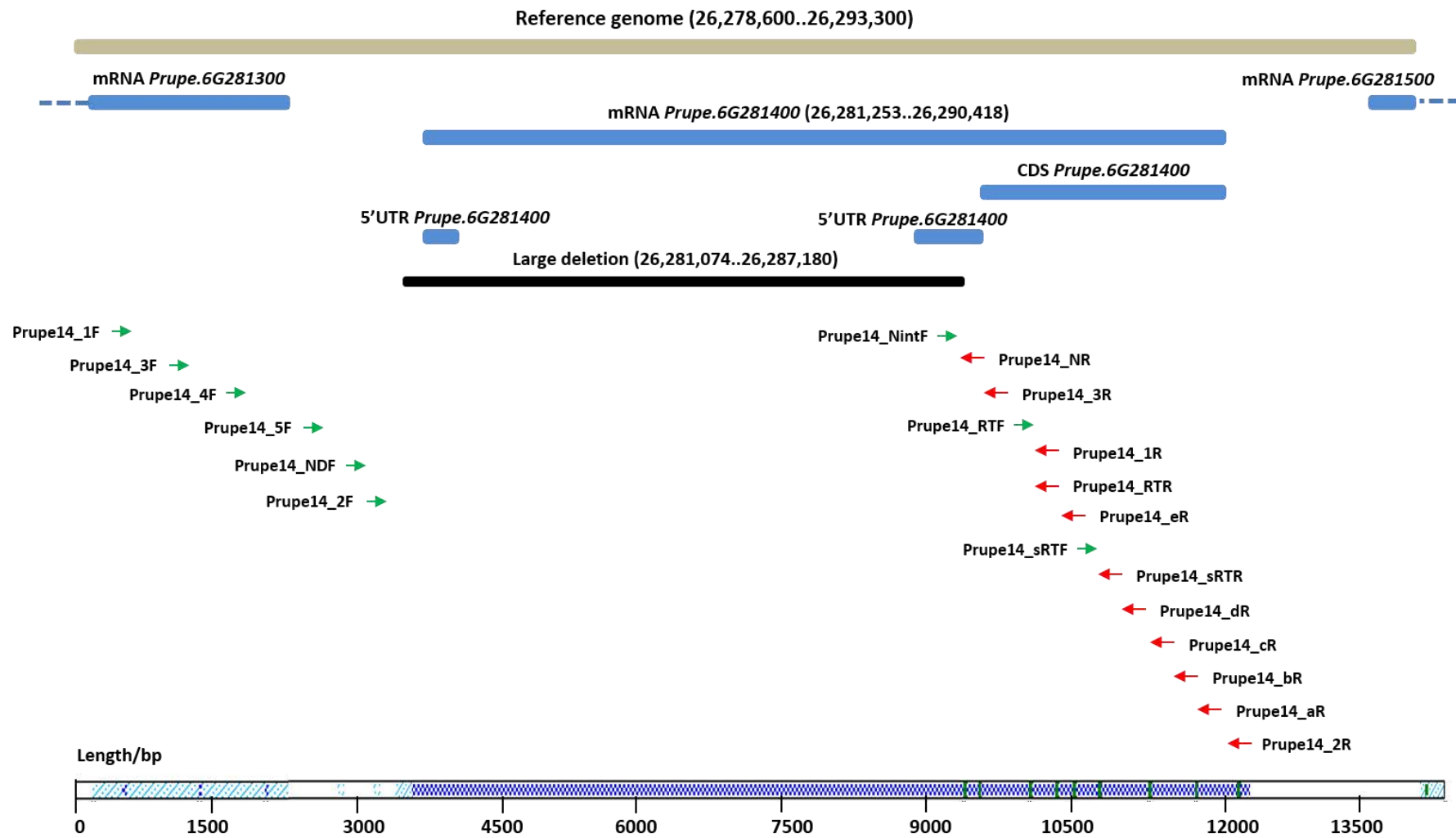
Supplementary Figure 2.S1 Overview of primers located during LRR-RL Kinases.

Supplementary Sequence 2.S1 Sequences of 2 Kb upstream and 3.5 Kb downstream deletion by Sanger sequencing. Marker "//" represents deletion breaking point.

>deletion-carrying allele in M14 (Pp06: 26,278,923..26,290,731)
AGAATCCTCTTATCTTCCATCTTCAGCCTCAATTTGGCCATAATATGTTTGGCTATGGCTCTCGCATTCCG
TCAGCAATCGGAAAGTTTACCAACTGGCATTGATGATTTCCTCACTTAGCGAGTTTCCTAACATCTCTTC
CTACTTAGTTACCCTTGATTTAAGCTTCAACAATTTGGAAAGACAAATGCCTGTGTCTATCTTCAATTTTC
GAGGCCTCGAATCACTTTGTCTTTCTTTCAACAATTTCACTGCCTTTCCATTCGATGGCCCTGAACAGCTG
ACAAAATCTTACAAACATTGACCTTTCGTACAATAGCTTGCTGAGTTTATACAATGGTACTAATTCATCA
TATTCCTCATTTCCTCAAATTGCTACATTGAATTTGGCTTCAAACAAATTCAGAACAATCCCAAATTTCTT
GAGAAATCAATCTGCATTATTTTCTTTGGACCTTTCATAAAACTGGATTGAAGGCAAGATACCTCGTTGG
ATTTGGAGTTTCAAATATTTTTATGACTTAAATCTCTCTTGTAACTCCTTAGGAACTCCAGAAGCTCCTTT
CTCTAAGCCTGATGTTTGTGAGCTTGACCTTCATTCAAACGAACTCCGGGGGAAACTCCCGATTATCCTA
CCGAATGCCTACTATCTTGATTACTCACAAATAATTTCAGCTCTATCATACCAACTGGCATTGGTGATTT
CCTCACTTCCGATACTTCCTTCCTATCTCTTTCAAGCAATAACATGCATGGCCACATTCCAGTATCGATAT
GCAATGGGGGTCTTCGGGTGCTTGATCTATCCAATAATTCTCTGAGTGGCATGATTCCGCAATGCTTGT
CAACAAAAGCAATTATTGGAGTACTTAATTTAAGGAAAAACAACCTTACTGGAACTATTTCTAATTTTGA
AATTCCAGAGTATTATGAATTAGACACTCTAAACCTGGGTGAAAACCATATAGAAGGTCAGTTTCCAAA
ATCTCTAACCAACTGCACAGTTTTACAGGTTTTAAACCTTGGAAAGAACCATCTAGCATATTCCTTTCCAT
GCTTGTTGAAGAATATATCCACTCTACGTGTCCTTGTTTTGCAGTCCAACAAATTCTACGGACGCACTGA
ATGTCCCAAGACCAATGGCACCTGGCCAATGCTTCAAATCATACCGAGAAGATCTTTGACAACATGGCG
GTCAATGATGGCTAACAAAGACGACTCCCTAACAGAGGTCAAATTCCTAGAATTTTCACATTAAGCAGG
GAGAGACGCGCGGGGTTGGTTTTTCTTTTGAGGATGGTATAACAGTTACCAGCAAAGGTTCAGAGATG
GATCTGCGAAAGATTCTATCTATCTTCACCTTGATTGACTTCTCATGCAAAATTTTCAGTGGAACAAAGG
AAATTGGAGAATTCAAATCACTATATGTCCTTAACTTGTCCGGAAATGCCTTGACAGGGGAAATCCCAT
CCTCCATTTGGTAACATGAGGGTACTCGAGTCCTTGGACCTGTCACAGAACAAACTGAGTGGGCAAATT
CCACCACAGTTGGCAAAGCTGACTTTCCTTTCATTCTTGAACCTCTCAAACAATCAACTGGTAGGCAGGA
TCCCAACCAGTACTCAGTTTTCCACATAAATCAAAGGGGGAGACGTTGAAGGGATTATGAAATTTATTT
TTTTCTTGTGTAAGTTTCAATAATTGGGATTTAAGCATATCAAAAGCCCAGGTTACAAATGGGACTTAA
ATTTTACTTTTATTATTTGTTATGAATGACCCAAATTGTTTTGTTTAAGAAACAGAACTTTTACTAACACG
GTCCTTATTTTCTAATTGCATTCCAAGAAGAAGAATAATATTTGAACATCCATTCATTTTAACAGAGGTT
ACGTAACAATATGATCCAAAACTACATGTAAAACTATACACGTACTTGCCAAGTTCACGGCTGGTTTTCT
TTGTTATAAATGGCTCAAATTCATCCCGAGGCCTGACCATGCAATTTCACAATTCAAACTTTATCAATATT
TTCATGAACTCTATATCTTCAGCTTTATCAATATTTTCATCTTCCAACCCAGAAATCCATCGACTTATTCAA
TAGTCAATAGTCAAAATTTTCATCTTCAAAGTCAACAATTCAAGCAAACCACAAATGTTGTGGAGACATT
TTCCACAAATT//TAATCTAAAGCCACCATATCCATCAATATTCCTCAGCGTCTCTATTTCCACGACTGCTG
AAGCCACATTTGCAAACTAATATCCTAGCCTATTCTGTGCTTCAACACAATTGCTGTCACATATCACAGC
CAGCAAGAGCTTCCAATGAGAATTCCTCTGTTTTCATGGCTTTTCTTAATCCTTATTCACTATGTTTCACT
CAGCGTCCATGTATTTGTGGTCTCTAGTCAATGTCCCAGCGACGATCAGCAATCTTTGTTGCTTCAATTG
AAGAACAGCCTCCAATTTGACTCTGCAAAATCTAACAAACTTAAGCAGTGGAAAAATGGCTCAGATTAC
TGCTCTTGGGAAGGTGTGTCTTGCAAAGATGGATGTGTTTCTCATCTCGACTTAAGCAGCGAGTCAATT
TCAGGAGGAGTTGATAATTCAAGTGCTCTTTTTGATCTGCAGTACATCGAGAACCTGAATTTGGCTTAC
AATAACTTCAACACTCAGATTCCATCCAAGTTCGACAGGCTGACCGGTTTGAGTTATCTGAACCTGTCAA
ATGCTGGCTTTGTGGGGCAGATTCCAATTGAGATTTCACACCTGGCAAGGTTGGTAACTCTTGATTTAT
CTACCTTTTACTTTCCCGGTCCACTGAAACTTGAGAATCCAAATTTGAATGTGCTCCTTGGCAACCTTTCT
GAGCTTATTGAACTTCATCTTGATGGTGTGAGCATATCAGATCAGGGAACTGAGTGGTGCGAAGTGAT
ATCGTCTTCACTGCCAAAGTTGAAGGTGTTGAGCTTGTCCACTTGTAATCTTTCAGGCCCTATTGATATT
TCATTACAGAAGCTTCACTCCCTCTCAGTTATCCGTTTAGAGAACAATAATTTGTCTGCTCAAGTTCCAG
AATTCTTCTCAAATTTCACAAATTTGACTTCGTTGCACCTCAGTAATTCTGGGTTAGATGGTACATTCCCA
AAGAAGATCTTCCAAGTACCTACACTGCAGACTATTGACTTGTCGGGTAATCAACAGCTTCAAGGTTCTT
TACCAGAATTTCCGAAGAATGGATCACTTCGTTCCTTGGTTCTAAGTGGGGCAAATTTTACGGGGTTTTT
TCCAAGCTCTATGGGCGACCTCAAAATGTTGTCCAGAATAGATGTTTCAAGTTGCAATTTCACTGGATCA

ACCCCAAGCTCAATGGAAAACCTTACACAATTGGTTTCTGTGGACCTCTCATGGAACAAGTTCAATGGT
TCTATTCCATTCTTCAGTATGGCCAAGAATCTGACCCTAATAAATCTTTCTTACAATCAGCTAACAGGTCA
GATTAATTCCTCTCATTGGGAAAACCTTACTAATCTGGTGAATCTCGACATGTGACACAATCTACTTGA
WGGAACTATTCCACCGTCTCTGTTTTCTCTTCCCGTGTTGCTGAAACTACAGCTTTCTAACAATCAACTCT
CTGGCCAGTTGCCTGTATTTGTTGGTATCTCTCTACTGGACACCCTTGATTTGAATAGCAATAAGTTGGA
AGGGCCTATACCAATGTCTWYYTTTAATCTCAAAGGGCTTAAGATTCTTTCACTTTCTTCAAACAACTTC
AGCGGCTCCTTTCCTCTTGAGCTTCTTCCACAACTGAAAAATCTTTCGAGTCTTGATCTTTCGTACAATAG
TTTGTCGATTAATTACAACACCCCCAATTCCTCTGTCACTTCCTTTCCTCAAATTACCACATTGAAATTGG
CATCTGTCAGGTTGAGAAGATTCCCAGATTTCTTAAGAGACCAATCCCACTTAAGCAATTTGGACCTTTC
ACAAAACCAGATTCGTGGAGAGATACCCAACTGGATTTGGAGGCTCAATAATCTTTCTCAACTAAATCT
TTCTTGCAACTCTCTGGAAACTCTAGAAGGTCCTCTCCTCAATGTTACTTCTAGTTTGTCTCTCCTTGACC
TTCATTCCAACCAGCTTAAGGGACAAATCCCACTTCTTTCACAATTTTCAGTTTATATAGATTATTCAAGA
AATAACTTTAACTCTAGCATACGGACTGACATTGGTGATTTCCTTTCTAATACTATATTCTTCTCTCTTTTA
AACAATAAATTTCAAGGAATCATTCCAGAATCTATATGCAATGCGTCAAATCTTCAGGTTCTTGATGTTT
CAAATAATTCTCTCAGTGGCTTGATTCCCAAGTGCTTGACTGCAATAAGTGGGACTCTTGCGGTACTGA
ATTTGAGAAGAAACAATCTTTCTGGCACTGTTTCTGGCACAATGTTAGAGGTTTTGAACCTTGGAAACA
ATCAAATAGCAGATACGTTTCCTTGCCTGTTGAAGAACATTTCCACCTTGTTCTTGTTTTGCGATCCAACA
AATTTTATGGACGCTTTGGATGCCCCAAGCCCCATGGAAACTGGTCAGTGCTTCAAATTGTTGACATAG
CACTCAACAATTTTAGTGGTGAAATACGAGGAAAATGCTTGAGAACCTGGAAGGCAATGATGGGTGAC
GACGATGATGTCATGTCAGAGCATAATCACCTTCGATTTCAGGTCCTTGAAGTAGAAGAGGTTTATTAT
CAAGATWCAGTTACAGTTATTAACAAAGGGCTAGAGATGCAGCTAGTCAAGATTCAACTGTCTTCACCT
CGATTGACCTCTCTTGCAACAAATTCACTGGATCAATACCTAAGGAAATGGGAGATCTCATATCACTCTA
TGTTCTCAACTTGTCTAGTAATGCTCTAACAGGTGAAATCCCCTCATCTATGGGTGACTTACAAGATGTT
GAGTCCTTGGACCTGTCAAATAACAAGCTGAGCGGGCAAATTCCACCACAGTTGGCAAAGCTGACTTTC
CTTTCATTCTTGAACCTCTCAAACAATCAACTGGTGGGCAGGATCCCAATCAGTACTCAGTTTTCAACAT
TTCCAAAAGCCTCCTTTACAGGTAACACAAGATTATCGGGGCCTCCTTTGACGGTCGACAACAAATCAG
TATCACCACCACCAACAGTGAATGGAAGCCCTCGAAATTGTGGACATCATCTTGAGATTAATTGGGGTA
TTATCAGTGTTGAGATTGGTTTTACAATTGGCTTTGGAGTTGCCATTGGCTCACTTGTGTTGTGCAAGAG
ATGGAGTAAATGGTATTACAAAGCTATGTATAACATCCTTCTCAAGATATTCCCTGAGCTGGAGGAAAG
AATTGGCATTCATCGAAGACATGTTAACATAAAATAAAGGTGGAGACGTTGAAATGATTATGAGGAAC
AATGCAACTGGACTCCATGTTTTGCCTAAGGTATTTTGAGATCAAAGAAATCCGAAGTGTGCTTTCTGG
TTTCATAGTTAATTTAATTGTGCTCTAGTTAGTCTTGGAATATCATGTATCTTTCATATTTTATTTTTTGTG
GCATGTAGAAAACATTACATAAATTCAAATATCAGCAAATCATTACTAACTACTTGTGGACAGACCGTCC
CCGCTGCGTTTGGTATTGCAGTGGGTAAAGCAAAAGACCTTCTATTTCTGTTTTTAAAGTAAAAGGCCC

Note: reference allele are not shown.

Supplementary Table 2.S2 BLAST analysis of *Prupe.6G281400* translated protein against the PLAZA protein sequence database.

Sequences alignments	Homologous blast	Orthologous blast	Score (Bits)	E Value	Species	Description (annoMine)
PPE_006G28830	HOM03D000009	ORTHO03D028407	1635	0.0	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G095720	HOM03D000009	ORTHO03D000539	1212	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD12G018290	HOM03D000009	ORTHO03D035040	1179	0.0	*Malus domestica*	Brassinosteroid LRR receptor kinase
MD04G012850	HOM03D000009	ORTHO03D000539	1168	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G095710	HOM03D000009	ORTHO03D000539	1162	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G282960	HOM03D000009	ORTHO03D000539	1153	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
FV6G10050	HOM03D000009	ORTHO03D018352	1132	0.0	*Fragaria vesca*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD12G018370	HOM03D000009	ORTHO03D000539	1123	0.0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
FV6G10040	HOM03D000009	ORTHO03D018352	1093	0.0	*Fragaria vesca*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD12G018180	HOM03D000009	ORTHO03D000539	1029	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G282930	HOM03D000009	ORTHO03D000539	1010	0.0	*Malus domestica*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
MD00G351530	HOM03D000009	ORTHO03D000539	1006	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
FV0G24790	HOM03D000009	ORTHO03D018352	998	0.0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G29520	HOM03D000009	ORTHO03D028407	998	0.0	*Prunus persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
FV6G10690	HOM03D000009	ORTHO03D018352	982	0.0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
MD05G014070	HOM03D000009	ORTHO03D000539	981	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD04G012840	HOM03D000009	ORTHO03D000539	960	0.0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
FV6G09780	HOM03D000009	ORTHO03D018352	948	0.0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
CS00098G00490	HOM03D000009	ORTHO03D001987	934	0.0	*Citrus sinensis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
FV6G27840	HOM03D000009	ORTHO03D092018	926	0.0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
MD08G006770	HOM03D000009	ORTHO03D028407	917	0.0	*Malus domestica*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
GR06G18410	HOM03D000009	ORTHO03D001987	917	0.0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
CP00120G00230	HOM03D000009	ORTHO03D001987	914	0.0	*Carica papaya*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
FV0G24860	HOM03D000009	ORTHO03D018352	910	0.0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
ME02042G00010	HOM03D000009	ORTHO03D001987	907	0.0	*Manihot esculenta*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G095690	HOM03D000009	ORTHO03D000539	907	0.0	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

CS01786G00010	HOM03D000009	ORTHO03D001987	905	0.0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G28810	HOM03D000009	ORTHO03D018352	904	0.0	*Prunus persica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
CP00120G00220	HOM03D000009	ORTHO03D001987	901	0.0	*Carica papaya*	receptor-like kinase protein THICK TASSEL DWARF1
TC0005G01160	HOM03D000009	ORTHO03D001987	900	0.0	*Theobroma cacao*	-
MD00G077620	HOM03D000009	ORTHO03D000539	895	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
RC29601G00090	HOM03D000009	ORTHO03D001987	893	0.0	*Ricinus communis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ST09G012330	HOM03D000009	ORTHO03D000539	890	0.0	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
TC0005G01150	HOM03D000009	ORTHO03D001987	887	0.0	*Theobroma cacao*	-
FV4G16710	HOM03D000009	ORTHO03D018352	886	0.0	*Fragaria vesca*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G109640	HOM03D000009	ORTHO03D000539	881	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
CS00098G00500	HOM03D000009	ORTHO03D001987	880	0.0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
ME04115G00010	HOM03D000009	ORTHO03D001987	879	0.0	*Manihot esculenta*	receptor-like kinase protein THICK TASSEL DWARF1
ME02042G00020	HOM03D000009	ORTHO03D001987	879	0.0	*Manihot esculenta*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
EG0001G04560	HOM03D000009	ORTHO03D000539	879	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
CS00098G00470	HOM03D000009	ORTHO03D001987	877	0.0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G088240	HOM03D000009	ORTHO03D023912	877	0.0	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PT16G12060	HOM03D000009	ORTHO03D001987	867	0.0	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017640	HOM03D000009	ORTHO03D000539	863	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
SL09G005080	HOM03D000009	ORTHO03D000539	863	0.0	*Solanum lycopersicum*	receptor-like kinase protein THICK TASSEL DWARF1
CM00011G00590	HOM03D000009	ORTHO03D000539	862	0.0	*Cucumis melo*	receptor-like kinase protein THICK TASSEL DWARF1
CL10G20240	HOM03D000009	ORTHO03D000539	861	0.0	*Citrullus lanatus*	receptor-like kinase protein THICK TASSEL DWARF1
GR02G25940	HOM03D000009	ORTHO03D001987	861	0.0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
CL10G20230	HOM03D000009	ORTHO03D000539	860	0.0	*Citrullus lanatus*	receptor-like kinase protein THICK TASSEL DWARF1
ST09G012340	HOM03D000009	ORTHO03D000539	857	0.0	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
GR12G11700	HOM03D000009	ORTHO03D001987	855	0.0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G095650	HOM03D000009	ORTHO03D000261	854	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
GM03G07240	HOM03D000009	ORTHO03D000539	850	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017720	HOM03D000009	ORTHO03D000539	848	0.0	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

EG0001G04500	HOM03D000009	ORTHO03D000539	840	0.0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
SL09G005090	HOM03D000009	ORTHO03D000539	839	0.0	*Solanum lycopersicum*	receptor-like kinase protein THICK TASSEL DWARF1
RC29601G00080	HOM03D000009	ORTHO03D000539	839	0.0	*Ricinus communis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G04530	HOM03D000009	ORTHO03D000539	838	0.0	*Eucalyptus grandis*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
MD00G351490	HOM03D000009	ORTHO03D001987	830	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
GR02G25950	HOM03D000009	ORTHO03D001987	826	0.0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
GM18G43510	HOM03D000009	ORTHO03D000539	826	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04360	HOM03D000009	ORTHO03D000539	821	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
GM01G31711	HOM03D000009	ORTHO03D000539	820	0.0	*Glycine max*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD08G006780	HOM03D000009	ORTHO03D000539	818	0.0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR02G26450	HOM03D000009	ORTHO03D001987	816	0.0	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G04380	HOM03D000009	ORTHO03D000539	815	0.0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0001G04550	HOM03D000009	ORTHO03D028407	814	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
CM00011G00580	HOM03D000009	ORTHO03D000539	811	0.0	*Cucumis melo*	receptor-like kinase protein THICK TASSEL DWARF1
GM18G43520	HOM03D000009	ORTHO03D000539	811	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
GR02G26230	HOM03D000009	ORTHO03D001987	809	0.0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
GM03G18170	HOM03D000009	ORTHO03D000539	794	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04430	HOM03D000009	ORTHO03D000539	791	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
PT16G12050	HOM03D000009	ORTHO03D001987	791	0.0	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
GM07G08770	HOM03D000009	ORTHO03D000539	787	0.0	*Glycine max*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MT4G017700	HOM03D000009	ORTHO03D000539	785	0.0	*Medicago truncatula*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
MD04G012900	HOM03D000009	ORTHO03D000261	778	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G013315	HOM03D000009	ORTHO03D000539	777	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017370	HOM03D000009	ORTHO03D000539	774	0.0	*Medicago truncatula*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
GM01G29570	HOM03D000009	ORTHO03D000539	773	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
GM18G43630	HOM03D000009	ORTHO03D000539	772	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
GM18G43621	HOM03D000009	ORTHO03D000539	771	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017730	HOM03D000009	ORTHO03D000539	767	0.0	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

MT4G417260	HOM03D000009	ORTHO03D000539	765	0.0	*Medicago truncatula*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G04420	HOM03D000009	ORTHO03D000539	759	0.0	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
BV5G09940	HOM03D000009	ORTHO03D000539	759	0.0	*Beta vulgaris*	Receptor-like protein kinase BRI1-like 3
LJ1G031850	HOM03D000009	ORTHO03D000539	759	0.0	*Lotus japonicus*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MT4G018940	HOM03D000009	ORTHO03D000539	756	0.0	*Medicago truncatula*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GM01G29030	HOM03D000009	ORTHO03D000539	756	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017280	HOM03D000009	ORTHO03D000539	754	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
GM01G29615	HOM03D000009	ORTHO03D000539	753	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
GM01G28936	HOM03D000009	ORTHO03D000539	752	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04390	HOM03D000009	ORTHO03D000539	749	0.0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0001G05420	HOM03D000009	ORTHO03D000539	747	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017600	HOM03D000009	ORTHO03D000539	744	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
MT5G046350	HOM03D000009	ORTHO03D000539	743	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G417270	HOM03D000009	ORTHO03D000539	741	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G05460	HOM03D000009	ORTHO03D000539	739	0.0	*Eucalyptus grandis*	Brassinosteroid LRR receptor kinase
MD00G452650	HOM03D000009	ORTHO03D000539	738	0.0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR12G11720	HOM03D000009	ORTHO03D001987	734	0.0	*Gossypium raimondii*	Brassinosteroid LRR receptor kinase
GM03G22050	HOM03D000009	ORTHO03D000539	734	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
CS01117G00010	HOM03D000009	ORTHO03D000539	734	0.0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017350	HOM03D000009	ORTHO03D000539	732	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04450	HOM03D000009	ORTHO03D000539	732	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G458530	HOM03D000009	ORTHO03D000261	731	0.0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G04520	HOM03D000009	ORTHO03D000539	729	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
LJ2G010060	HOM03D000009	ORTHO03D000539	727	0.0	*Lotus japonicus*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G282920	HOM03D000009	ORTHO03D000261	721	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
GM01G29580	HOM03D000009	ORTHO03D000539	711	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
CS01062G00010	HOM03D000009	ORTHO03D000539	707	0.0	*Citrus sinensis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G088230	HOM03D000009	ORTHO03D000261	702	0.0	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

EG0001G05440	HOM03D000009	ORTHO03D000539	694	0.0	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
CS00019G01260	HOM03D000009	ORTHO03D001987	693	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G198270	HOM03D000009	ORTHO03D000261	690	0.0	*Vitis vinifera*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR02G26470	HOM03D000009	ORTHO03D000261	670	0.0	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G04470	HOM03D000009	ORTHO03D000261	659	0.0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
VV08G06950	HOM03D000009	ORTHO03D036633	653	0.0	*Vitis vinifera*	receptor-like kinase protein FLORAL ORGAN NUMBER1
PPE_008G06340	HOM03D000009	ORTHO03D000261	653	0.0	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G05510	HOM03D000009	ORTHO03D000539	650	0.0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_007G11060	HOM03D000009	ORTHO03D000261	645	0.0	*Prunus persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PT01G12840	HOM03D000009	ORTHO03D001987	638	0.0	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G019030	HOM03D000009	ORTHO03D000539	634	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G28840	HOM03D000009	ORTHO03D000261	632	0.0	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G019010	HOM03D000009	ORTHO03D000539	631	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
GM03G06804	HOM03D000009	ORTHO03D036633	630	0.0	*Glycine max*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
GM18G43500	HOM03D000009	ORTHO03D000539	630	0.0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
ST10G018600	HOM03D000009	ORTHO03D023912	629	0.0	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ATR_00029G03770	HOM03D000009	ORTHO03D000539	618	0.0	*Amborella trichopoda*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ATR_00029G03900	HOM03D000009	ORTHO03D000539	616	0.0	*Amborella trichopoda*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GM07G18640	HOM03D000009	ORTHO03D000539	615	0.0	*Glycine max*	receptor-like kinase protein FLORAL ORGAN NUMBER1
FV7G08990	HOM03D000009	ORTHO03D000261	613	0.0	*Fragaria vesca*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G485250	HOM03D000009	ORTHO03D000261	610	0.0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017710	HOM03D000009	ORTHO03D000539	609	0.0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_008G06330	HOM03D000009	ORTHO03D000261	602	0.0	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
ST10G018890	HOM03D000009	ORTHO03D023912	597	0.0	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0001G04340	HOM03D000009	ORTHO03D000539	593	0.0	*Eucalyptus grandis*	Brassinosteroid LRR receptor kinase
LJ2G010030	HOM03D000009	ORTHO03D000539	577	0.0	*Lotus japonicus*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0001G05490	HOM03D000009	ORTHO03D000261	576	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04170	HOM03D000009	ORTHO03D000261	573	0.0	*Eucalyptus grandis*	Receptor-like protein kinase BRI1-like 3

MD15G019820	HOM03D000009	ORTHO03D000261	563	0.0	*Malus domestica*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
EG0001G05480	HOM03D000009	ORTHO03D000261	558	0.0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
SL01G098690	HOM03D000009	ORTHO03D000261	557	0.0	*Solanum lycopersicum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD12G018150	HOM03D000009	ORTHO03D000261	540	0.0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
PPE_006G28780	HOM03D000009	ORTHO03D035040	545	2,00E-180	*Prunus persica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ST01G036380	HOM03D000009	ORTHO03D001987	550	8,00E-179	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
ST01G036370	HOM03D000009	ORTHO03D000261	548	5,00E-178	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MT2G078260	HOM03D000009	ORTHO03D000261	541	1,00E-175	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_006G28800	HOM03D000009	ORTHO03D000261	533	5,00E-175	*Prunus persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
SL01G098680	HOM03D000009	ORTHO03D000261	531	3,00E-172	*Solanum lycopersicum*	receptor-like kinase protein THICK TASSEL DWARF1
PT01G43770	HOM03D000009	ORTHO03D001987	528	2,00E-169	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
ZM05G26420	HOM03D000009	ORTHO03D049415	526	5,00E-168	*Zea mays*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G279640	HOM03D000009	ORTHO03D000261	519	2,00E-167	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G05430	HOM03D000009	ORTHO03D000261	503	6,00E-162	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
LJ2G010310	HOM03D000009	ORTHO03D000261	514	4,00E-161	*Lotus japonicus*	receptor-like kinase protein FLORAL ORGAN NUMBER1
BV0G79990	HOM03D000009	ORTHO03D000261	500	7,00E-159	*Beta vulgaris*	receptor-like kinase protein FLORAL ORGAN NUMBER1
FV6G27830	HOM03D000009	ORTHO03D000261	491	9,00E-158	*Fragaria vesca*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS12G12120	HOM03D000009	ORTHO03D000261	492	7,00E-156	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS04G40440	HOM03D000009	ORTHO03D000261	493	2,00E-155	*Oryza sativa ssp. Japonica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR05G04440	HOM03D000009	ORTHO03D000261	494	4,00E-155	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
TC0004G03690	HOM03D000009	ORTHO03D000261	493	4,00E-154	*Theobroma cacao*	-
GR02G26440	HOM03D000009	ORTHO03D000261	481	7,00E-154	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G018910	HOM03D000009	ORTHO03D000261	482	2,00E-152	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
ST10G018860	HOM03D000009	ORTHO03D000261	476	2,00E-151	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
OS12G11720	HOM03D000009	ORTHO03D000261	481	2,00E-151	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G12130	HOM03D000009	ORTHO03D000261	478	2,00E-150	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MT4G017260	HOM03D000009	ORTHO03D000261	473	6,00E-150	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
CS01062G00050	HOM03D000009	ORTHO03D000261	464	1,00E-149	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1

OS12G11370	HOM03D000009	ORTHO03D000261	476	2,00E-149	*Oryza sativa ssp. Japonica*	receptor-like kinase protein THICK TASSEL DWARF1
OS12G12010	HOM03D000009	ORTHO03D000261	474	5,00E-149	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ST09G012370	HOM03D000009	ORTHO03D000261	462	1,00E-148	*Solanum tuberosum*	Brassinosteroid LRR receptor kinase
MT4G017490	HOM03D000009	ORTHO03D000261	467	4,00E-148	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
SL10G076500	HOM03D000009	ORTHO03D000261	464	3,00E-147	*Solanum lycopersicum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ST10G018880	HOM03D000009	ORTHO03D000261	464	4,00E-147	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR05G02630	HOM03D000009	ORTHO03D000261	466	1,00E-145	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
CS00376G00030	HOM03D000009	ORTHO03D000261	457	2,00E-143	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
GR05G02940	HOM03D000009	ORTHO03D000261	460	3,00E-143	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
GM07G18590	HOM03D000009	ORTHO03D000261	453	1,00E-142	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017780	HOM03D000009	ORTHO03D000261	451	2,00E-141	*Medicago truncatula*	S-cell enriched with leucine-rich repeat-containing protein slrA
OS12G11680	HOM03D000009	ORTHO03D000261	454	4,00E-141	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G10870	HOM03D000009	ORTHO03D000261	449	1,00E-139	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G11930	HOM03D000009	ORTHO03D000261	449	3,00E-139	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
VV08G06930	HOM03D000009	ORTHO03D057804	458	3,00E-139	*Vitis vinifera*	receptor-like kinase protein THICK TASSEL DWARF1
OS01G06520	HOM03D000009	ORTHO03D000261	446	4,00E-138	*Oryza sativa ssp. Japonica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ME08617G00010	HOM03D000009	ORTHO03D000261	442	1,00E-137	*Manihot esculenta*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ME04305G00130	HOM03D000009	ORTHO03D000261	438	5,00E-136	*Manihot esculenta*	receptor-like kinase protein THICK TASSEL DWARF1
GR02G26240	HOM03D000009	ORTHO03D000261	432	2,00E-135	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ST12G032880	HOM03D000009	ORTHO03D000980	438	2,00E-135	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR09G44800	HOM03D000009	ORTHO03D000261	437	2,00E-134	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
TC0010G16450	HOM03D000009	ORTHO03D012248	434	4,00E-134	*Theobroma cacao*	-
OS01G06900	HOM03D000009	ORTHO03D000261	435	4,00E-134	*Oryza sativa ssp. Japonica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR05G02640	HOM03D000009	ORTHO03D000261	432	7,00E-134	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G11500	HOM03D000009	ORTHO03D000261	433	9,00E-134	*Oryza sativa ssp. Japonica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR09G44790	HOM03D000009	ORTHO03D012248	431	8,00E-133	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS01G06920	HOM03D000009	ORTHO03D000261	431	1,00E-132	*Oryza sativa ssp. Japonica*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
MT4G018920	HOM03D000009	ORTHO03D000261	418	3,00E-132	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1

GR09G44810	HOM03D000009	ORTHO03D012248	431	4,00E-132	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
TC0026G00080	HOM03D000009	ORTHO03D012248	426	6,00E-131	*Theobroma cacao*	-
OS01G06730	HOM03D000009	ORTHO03D000261	425	2,00E-130	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0011G05350	HOM03D000009	ORTHO03D000261	424	2,00E-130	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G351510	HOM03D000009	ORTHO03D000261	408	4,00E-130	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR09G44820	HOM03D000009	ORTHO03D012248	424	5,00E-130	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
SL07G005150	HOM03D000009	ORTHO03D028741	424	8,00E-130	*Solanum lycopersicum*	repeat receptor protein kinase EXS
TC0007G01810	HOM03D000009	ORTHO03D000261	424	8,00E-130	*Theobroma cacao*	-
OS01G06670	HOM03D000009	ORTHO03D000261	420	3,00E-128	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
PT01G38910	HOM03D000009	ORTHO03D094790	416	1,00E-127	*Populus trichocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G201230	HOM03D000009	ORTHO03D000261	406	2,00E-127	*Malus domestica*	repeat receptor protein kinase EXS
PT16G12690	HOM03D000009	ORTHO03D000261	416	3,00E-127	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
ST09G011940	HOM03D000009	ORTHO03D000261	407	3,00E-127	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
TC0007G01830	HOM03D000009	ORTHO03D088047	415	2,00E-126	*Theobroma cacao*	-
RC29848G00700	HOM03D000009	ORTHO03D000261	412	4,00E-126	*Ricinus communis*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
GR09G36770	HOM03D000009	ORTHO03D012248	412	4,00E-126	*Gossypium raimondii*	Brassinosteroid LRR receptor kinase
ST12G032280	HOM03D000009	ORTHO03D000980	410	1,00E-125	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR09G35460	HOM03D000009	ORTHO03D012248	411	2,00E-125	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G11860	HOM03D000009	ORTHO03D000261	411	2,00E-125	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0007G07160	HOM03D000009	ORTHO03D000261	410	3,00E-125	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR11G15540	HOM03D000009	ORTHO03D000261	410	1,00E-124	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR06G05630	HOM03D000009	ORTHO03D000261	408	1,00E-124	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
TC0026G00040	HOM03D000009	ORTHO03D012248	409	2,00E-124	*Theobroma cacao*	-
ZM03G04280	HOM03D000009	ORTHO03D000261	409	3,00E-124	*Zea mays*	receptor-like kinase protein THICK TASSEL DWARF1
MT5G086530	HOM03D000009	ORTHO03D014880	408	5,00E-124	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0011G05420	HOM03D000009	ORTHO03D000261	405	2,00E-123	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
SL12G100030	HOM03D000009	ORTHO03D000980	403	3,00E-123	*Solanum lycopersicum*	receptor-like kinase protein THICK TASSEL DWARF1
ZM03G04340	HOM03D000009	ORTHO03D000261	406	8,00E-123	*Zea mays*	receptor-like kinase protein THICK TASSEL DWARF1

EG0005G19670	HOM03D000009	ORTHO03D007020	402	8,00E-123	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0005G11440	HOM03D000009	ORTHO03D007020	401	3,00E-122	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
VV08G06980	HOM03D000009	ORTHO03D000261	395	5,00E-122	*Vitis vinifera*	S-cell enriched with leucine-rich repeat-containing protein slrA
EG0011G05400	HOM03D000009	ORTHO03D000261	401	1,00E-121	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
TC0007G14800	HOM03D000009	ORTHO03D000980	416	1,00E-121	*Theobroma cacao*	-
RC29848G00730	HOM03D000009	ORTHO03D000261	398	3,00E-121	*Ricinus communis*	Receptor-like protein kinase BRI1-like 3
SL12G100010	HOM03D000009	ORTHO03D000980	401	3,00E-121	*Solanum lycopersicum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR09G44860	HOM03D000009	ORTHO03D012248	399	2,00E-120	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
GR09G44440	HOM03D000009	ORTHO03D012248	397	2,00E-120	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MT5G087320	HOM03D000009	ORTHO03D014880	397	6,00E-120	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0011G05380	HOM03D000009	ORTHO03D000261	395	1,00E-119	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
PPE_003G08160	HOM03D000009	ORTHO03D015309	394	1,00E-119	*Prunus persica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0002G12380	HOM03D000009	ORTHO03D007020	393	1,00E-119	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G485210	HOM03D000009	ORTHO03D000261	377	2,00E-119	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
TC0007G10730	HOM03D000009	ORTHO03D012248	392	3,00E-118	*Theobroma cacao*	-
MT5G096340	HOM03D000009	ORTHO03D000980	392	6,00E-118	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PT12G02760	HOM03D000009	ORTHO03D082383	391	8,00E-118	*Populus trichocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0002G05880	HOM03D000009	ORTHO03D007020	389	1,00E-117	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PT00G11860	HOM03D000009	ORTHO03D000261	386	6,00E-117	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
GM03G07400	HOM03D000009	ORTHO03D000261	383	9,00E-117	*Glycine max*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
TC0007G14860	HOM03D000009	ORTHO03D000980	392	2,00E-116	*Theobroma cacao*	-
GR09G44890	HOM03D000009	ORTHO03D000980	389	2,00E-116	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0011G05390	HOM03D000009	ORTHO03D048476	385	2,00E-116	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GM16G28480	HOM03D000009	ORTHO03D000980	387	2,00E-116	*Glycine max*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0009G04410	HOM03D000009	ORTHO03D007020	388	5,00E-116	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
BR09G09110	HOM03D000009	ORTHO03D035676	384	1,00E-115	*Brassica rapa*	receptor-like kinase protein FLORAL ORGAN NUMBER1
PT00G14580	HOM03D000009	ORTHO03D084117	387	2,00E-115	*Populus trichocarpa*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GM18G43490	HOM03D000009	ORTHO03D000261	381	3,00E-115	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1

Supplementary Table 2.S3 BLAST analysis of *Prupe.6G281500* translated protein against the PLAZA protein sequence database.

Sequences aligments	Homologous blast	Orthologous blast	Score (Bits)	E Value	Species	Description (annoMine)
MD04G012850	HOM03D000009	ORTHO03D000539	1305	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
FV6G10040	HOM03D000009	ORTHO03D018352	1274	0	*Fragaria vesca*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G095720	HOM03D000009	ORTHO03D000539	1187	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G28830	HOM03D000009	ORTHO03D028407	1184	0	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
MD12G018370	HOM03D000009	ORTHO03D000539	1179	0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
FV6G10050	HOM03D000009	ORTHO03D018352	1144	0	*Fragaria vesca*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G351490	HOM03D000009	ORTHO03D001987	1117	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD12G018290	HOM03D000009	ORTHO03D035040	1117	0	*Malus domestica*	Brassinosteroid LRR receptor kinase
MD00G095710	HOM03D000009	ORTHO03D000539	1097	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G282960	HOM03D000009	ORTHO03D000539	1085	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD12G018180	HOM03D000009	ORTHO03D000539	1085	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G351530	HOM03D000009	ORTHO03D000539	1046	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
GR06G18410	HOM03D000009	ORTHO03D001987	1043	0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G28840	HOM03D000009	ORTHO03D000261	1039	0	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
CS00098G00500	HOM03D000009	ORTHO03D001987	1021	0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04560	HOM03D000009	ORTHO03D000539	1021	0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
ME04115G00010	HOM03D000009	ORTHO03D001987	1017	0	*Manihot esculenta*	receptor-like kinase protein THICK TASSEL DWARF1
CP00120G00230	HOM03D000009	ORTHO03D001987	1008	0	*Carica papaya*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G095690	HOM03D000009	ORTHO03D000539	981	0	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
CL10G20240	HOM03D000009	ORTHO03D000539	979	0	*Citrullus lanatus*	receptor-like kinase protein THICK TASSEL DWARF1
PT16G12060	HOM03D000009	ORTHO03D001987	979	0	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G282930	HOM03D000009	ORTHO03D000539	977	0	*Malus domestica*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
FV0G24790	HOM03D000009	ORTHO03D018352	971	0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G29520	HOM03D000009	ORTHO03D028407	966	0	*Prunus persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GM18G43510	HOM03D000009	ORTHO03D000539	962	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
FV0G24860	HOM03D000009	ORTHO03D018352	951	0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1

MT4G017640	HOM03D000009	ORTHO03D000539	950	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
CS00098G00490	HOM03D000009	ORTHO03D001987	948	0	*Citrus sinensis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD05G014070	HOM03D000009	ORTHO03D000539	944	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
ST09G012330	HOM03D000009	ORTHO03D000539	941	0	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
CS01786G00010	HOM03D000009	ORTHO03D001987	939	0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
MD08G006770	HOM03D000009	ORTHO03D028407	936	0	*Malus domestica*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
TC0005G01160	HOM03D000009	ORTHO03D001987	932	0	*Theobroma cacao*	-
FV6G09780	HOM03D000009	ORTHO03D018352	931	0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
LJ1G031850	HOM03D000009	ORTHO03D000539	920	0	*Lotus japonicus*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ME02042G00010	HOM03D000009	ORTHO03D001987	918	0	*Manihot esculenta*	receptor-like kinase protein FLORAL ORGAN NUMBER1
TC0005G01150	HOM03D000009	ORTHO03D001987	915	0	*Theobroma cacao*	-
MD04G012840	HOM03D000009	ORTHO03D000539	912	0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
RC29601G00080	HOM03D000009	ORTHO03D000539	910	0	*Ricinus communis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
CS00098G00470	HOM03D000009	ORTHO03D001987	908	0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
FV6G10690	HOM03D000009	ORTHO03D018352	905	0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
CP00120G00220	HOM03D000009	ORTHO03D001987	904	0	*Carica papaya*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G198270	HOM03D000009	ORTHO03D000261	903	0	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
SL09G005080	HOM03D000009	ORTHO03D000539	903	0	*Solanum lycopersicum*	receptor-like kinase protein THICK TASSEL DWARF1
CL10G20230	HOM03D000009	ORTHO03D000539	902	0	*Citrullus lanatus*	receptor-like kinase protein THICK TASSEL DWARF1
FV6G27840	HOM03D000009	ORTHO03D092018	902	0	*Fragaria vesca*	receptor-like kinase protein THICK TASSEL DWARF1
GR02G25940	HOM03D000009	ORTHO03D001987	899	0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G109640	HOM03D000009	ORTHO03D000539	899	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
RC29601G00090	HOM03D000009	ORTHO03D001987	898	0	*Ricinus communis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
CM00011G00590	HOM03D000009	ORTHO03D000539	895	0	*Cucumis melo*	receptor-like kinase protein THICK TASSEL DWARF1
MD08G006780	HOM03D000009	ORTHO03D000539	890	0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ME02042G00020	HOM03D000009	ORTHO03D001987	890	0	*Manihot esculenta*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
PPE_006G28810	HOM03D000009	ORTHO03D018352	890	0	*Prunus persica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR12G11720	HOM03D000009	ORTHO03D001987	885	0	*Gossypium raimondii*	Brassinosteroid LRR receptor kinase
EG0001G04500	HOM03D000009	ORTHO03D000539	882	0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

CM00011G00580	HOM03D000009	ORTHO03D000539	881	0	*Cucumis melo*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04380	HOM03D000009	ORTHO03D000539	876	0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0001G04530	HOM03D000009	ORTHO03D000539	875	0	*Eucalyptus grandis*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
EG0001G04550	HOM03D000009	ORTHO03D028407	875	0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
ST09G012340	HOM03D000009	ORTHO03D000539	875	0	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04360	HOM03D000009	ORTHO03D000539	872	0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
GR02G26450	HOM03D000009	ORTHO03D001987	862	0	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR02G26230	HOM03D000009	ORTHO03D001987	860	0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
BV5G09940	HOM03D000009	ORTHO03D000539	858	0	*Beta vulgaris*	Receptor-like protein kinase BRI1-like 3
MT4G017720	HOM03D000009	ORTHO03D000539	851	0	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PT16G12050	HOM03D000009	ORTHO03D001987	845	0	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
GM03G07240	HOM03D000009	ORTHO03D000539	844	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
SL09G005090	HOM03D000009	ORTHO03D000539	844	0	*Solanum lycopersicum*	receptor-like kinase protein THICK TASSEL DWARF1
GR12G11700	HOM03D000009	ORTHO03D001987	843	0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04420	HOM03D000009	ORTHO03D000539	840	0	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR02G25950	HOM03D000009	ORTHO03D001987	840	0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
GM07G18640	HOM03D000009	ORTHO03D000539	838	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04430	HOM03D000009	ORTHO03D000539	837	0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G077620	HOM03D000009	ORTHO03D000539	833	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD04G012900	HOM03D000009	ORTHO03D000261	829	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G095650	HOM03D000009	ORTHO03D000261	828	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04390	HOM03D000009	ORTHO03D000539	823	0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GM01G31711	HOM03D000009	ORTHO03D000539	818	0	*Glycine max*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GM18G43520	HOM03D000009	ORTHO03D000539	818	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
GM03G18170	HOM03D000009	ORTHO03D000539	813	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04520	HOM03D000009	ORTHO03D000539	811	0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G088240	HOM03D000009	ORTHO03D023912	811	0	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
FV4G16710	HOM03D000009	ORTHO03D018352	806	0	*Fragaria vesca*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MT4G017730	HOM03D000009	ORTHO03D000539	806	0	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

EG0001G05420	HOM03D000009	ORTHO03D000539	795	0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017700	HOM03D000009	ORTHO03D000539	793	0	*Medicago truncatula*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
GM07G08770	HOM03D000009	ORTHO03D000539	791	0	*Glycine max*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G05460	HOM03D000009	ORTHO03D000539	781	0	*Eucalyptus grandis*	Brassinosteroid LRR receptor kinase
GM18G43621	HOM03D000009	ORTHO03D000539	771	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04450	HOM03D000009	ORTHO03D000539	769	0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G013315	HOM03D000009	ORTHO03D000539	762	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G417260	HOM03D000009	ORTHO03D000539	762	0	*Medicago truncatula*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G05440	HOM03D000009	ORTHO03D000539	760	0	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GM01G29030	HOM03D000009	ORTHO03D000539	760	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017370	HOM03D000009	ORTHO03D000539	758	0	*Medicago truncatula*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
GM01G28936	HOM03D000009	ORTHO03D000539	756	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
GM03G22050	HOM03D000009	ORTHO03D000539	754	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
GM18G43630	HOM03D000009	ORTHO03D000539	754	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
CS01117G00010	HOM03D000009	ORTHO03D000539	753	0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
GM01G29570	HOM03D000009	ORTHO03D000539	751	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
GM01G29615	HOM03D000009	ORTHO03D000539	751	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017350	HOM03D000009	ORTHO03D000539	750	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017280	HOM03D000009	ORTHO03D000539	740	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
CS01062G00010	HOM03D000009	ORTHO03D000539	736	0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
CS00019G01260	HOM03D000009	ORTHO03D001987	731	0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G417270	HOM03D000009	ORTHO03D000539	723	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
MT5G046350	HOM03D000009	ORTHO03D000539	715	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
LJ2G010060	HOM03D000009	ORTHO03D000539	714	0	*Lotus japonicus*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G05510	HOM03D000009	ORTHO03D000539	709	0	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GM01G29580	HOM03D000009	ORTHO03D000539	708	0	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G04340	HOM03D000009	ORTHO03D000539	698	0	*Eucalyptus grandis*	Brassinosteroid LRR receptor kinase
MT4G017600	HOM03D000009	ORTHO03D000539	695	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
ATR_00029G03770	HOM03D000009	ORTHO03D000539	694	0	*Amborella trichopoda*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

MT4G017710	HOM03D000009	ORTHO03D000539	693	0	*Medicago truncatula*	Brassinosteroid LRR receptor kinase
GM18G43500	HOM03D000009	ORTHO03D000539	689	0	*Glycine max*	Brassinosteroid LRR receptor kinase
MD00G452650	HOM03D000009	ORTHO03D000539	685	0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ATR_00029G03900	HOM03D000009	ORTHO03D000539	679	0	*Amborella trichopoda*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G088230	HOM03D000009	ORTHO03D000261	678	0	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MT4G018940	HOM03D000009	ORTHO03D000539	677	0	*Medicago truncatula*	receptor-like kinase protein FLORAL ORGAN NUMBER1
PT01G12840	HOM03D000009	ORTHO03D001987	674	0	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
ME04305G00130	HOM03D000009	ORTHO03D000261	671	0	*Manihot esculenta*	receptor-like kinase protein THICK TASSEL DWARF1
ST01G036380	HOM03D000009	ORTHO03D001987	656	0	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
SL01G098690	HOM03D000009	ORTHO03D000261	655	0	*Solanum lycopersicum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ST01G036370	HOM03D000009	ORTHO03D000261	650	0	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR02G26470	HOM03D000009	ORTHO03D000261	644	0	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MT4G019010	HOM03D000009	ORTHO03D000539	643	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
FV7G08990	HOM03D000009	ORTHO03D000261	640	0	*Fragaria vesca*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G04470	HOM03D000009	ORTHO03D000261	639	0	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
LJ2G010030	HOM03D000009	ORTHO03D000539	637	0	*Lotus japonicus*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MT2G078260	HOM03D000009	ORTHO03D000261	635	0	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD00G458530	HOM03D000009	ORTHO03D000261	632	0	*Malus domestica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G282920	HOM03D000009	ORTHO03D000261	626	0	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
VV08G06950	HOM03D000009	ORTHO03D036633	624	0	*Vitis vinifera*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ST10G018890	HOM03D000009	ORTHO03D023912	623	0	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_007G11060	HOM03D000009	ORTHO03D000261	620	0	*Prunus persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GM03G06804	HOM03D000009	ORTHO03D036633	613	0	*Glycine max*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
LJ2G010310	HOM03D000009	ORTHO03D000261	602	0	*Lotus japonicus*	receptor-like kinase protein THICK TASSEL DWARF1
PT01G43770	HOM03D000009	ORTHO03D001987	595	0	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_008G06340	HOM03D000009	ORTHO03D000261	588	0	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
ZM05G26420	HOM03D000009	ORTHO03D049415	581	0	*Zea mays*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR02G26440	HOM03D000009	ORTHO03D000261	580	0	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
EG0001G05490	HOM03D000009	ORTHO03D000261	578	0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1

MT4G019030	HOM03D000009	ORTHO03D000539	574	0	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
ST09G012370	HOM03D000009	ORTHO03D000261	567	0	*Solanum tuberosum*	Brassinosteroid LRR receptor kinase
OS12G10870	HOM03D000009	ORTHO03D000261	557	2.00E-180	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G05480	HOM03D000009	ORTHO03D000261	554	0	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
ST10G018600	HOM03D000009	ORTHO03D023912	554	2.00E-179	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS04G40440	HOM03D000009	ORTHO03D000261	554	2.00E-178	*Oryza sativa ssp. Japonica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
SL10G076500	HOM03D000009	ORTHO03D000261	553	0	*Solanum lycopersicum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS12G12010	HOM03D000009	ORTHO03D000261	553	4.00E-179	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
SL01G098680	HOM03D000009	ORTHO03D000261	551	1.00E-179	*Solanum lycopersicum*	receptor-like kinase protein THICK TASSEL DWARF1
CS01062G00050	HOM03D000009	ORTHO03D000261	548	0	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
OS12G11720	HOM03D000009	ORTHO03D000261	548	6.00E-177	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G11370	HOM03D000009	ORTHO03D000261	547	2.00E-176	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G12120	HOM03D000009	ORTHO03D000261	546	4.00E-176	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G12130	HOM03D000009	ORTHO03D000261	543	7.00E-175	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
BV0G79990	HOM03D000009	ORTHO03D000261	541	2.00E-174	*Beta vulgaris*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS12G11860	HOM03D000009	ORTHO03D000261	541	3.00E-174	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ST10G018880	HOM03D000009	ORTHO03D000261	540	8.00E-176	*Solanum tuberosum*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS12G11680	HOM03D000009	ORTHO03D000261	533	9.00E-171	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR02G26240	HOM03D000009	ORTHO03D000261	527	1.00E-171	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR05G04440	HOM03D000009	ORTHO03D000261	524	3.00E-166	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS12G11930	HOM03D000009	ORTHO03D000261	521	2.00E-166	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MT4G017490	HOM03D000009	ORTHO03D000261	520	4.00E-168	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G017780	HOM03D000009	ORTHO03D000261	518	9.00E-167	*Medicago truncatula*	S-cell enriched with leucine-rich repeat-containing protein slrA
CS00376G00030	HOM03D000009	ORTHO03D000261	518	2.00E-166	*Citrus sinensis*	receptor-like kinase protein THICK TASSEL DWARF1
ME08617G00010	HOM03D000009	ORTHO03D000261	517	4.00E-166	*Manihot esculenta*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS12G11500	HOM03D000009	ORTHO03D000261	515	1.00E-164	*Oryza sativa ssp. Japonica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
TC0004G03690	HOM03D000009	ORTHO03D000261	514	8.00E-162	*Theobroma cacao*	-
ST09G011940	HOM03D000009	ORTHO03D000261	513	3.00E-167	*Solanum tuberosum*	receptor-like kinase protein THICK TASSEL DWARF1
PPE_006G28800	HOM03D000009	ORTHO03D000261	513	8.00E-167	*Prunus persica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

PPE_006G28780	HOM03D000009	ORTHO03D035040	511	8.00E-167	*Prunus persica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0001G05430	HOM03D000009	ORTHO03D000261	511	5.00E-165	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
MT4G018910	HOM03D000009	ORTHO03D000261	508	9.00E-162	*Medicago truncatula*	receptor-like kinase protein THICK TASSEL DWARF1
OS01G06900	HOM03D000009	ORTHO03D000261	508	2.00E-161	*Oryza sativa ssp. Japonica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_008G06330	HOM03D000009	ORTHO03D000261	505	2.00E-162	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
OS01G06920	HOM03D000009	ORTHO03D000261	505	1.00E-160	*Oryza sativa ssp. Japonica*	inactive leucine-rich repeat receptor-like protein kinase At3g28040
GM18G43490	HOM03D000009	ORTHO03D000261	503	7.00E-161	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G279640	HOM03D000009	ORTHO03D000261	502	1.00E-160	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
PT00G11860	HOM03D000009	ORTHO03D000261	501	7.00E-160	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
PT16G12690	HOM03D000009	ORTHO03D000261	497	1.00E-157	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
OS01G06520	HOM03D000009	ORTHO03D000261	496	4.00E-157	*Oryza sativa ssp. Japonica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR05G02630	HOM03D000009	ORTHO03D000261	496	1.00E-156	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
MD00G485250	HOM03D000009	ORTHO03D000261	493	3.00E-162	*Malus domestica*	receptor-like kinase protein THICK TASSEL DWARF1
TC0010G16450	HOM03D000009	ORTHO03D012248	490	3.00E-155	*Theobroma cacao*	-
PT01G38910	HOM03D000009	ORTHO03D094790	487	4.00E-154	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1
OS01G06730	HOM03D000009	ORTHO03D000261	486	3.00E-153	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
RC29848G00700	HOM03D000009	ORTHO03D000261	483	2.00E-152	*Ricinus communis*	inactive leucine-rich repeat receptor-like protein kinase At2g25790
GR09G35460	HOM03D000009	ORTHO03D012248	482	8.00E-152	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0005G11580	HOM03D000009	ORTHO03D007020	481	1.00E-152	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
EG0005G11440	HOM03D000009	ORTHO03D007020	480	8.00E-152	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ZM03G04280	HOM03D000009	ORTHO03D000261	479	1.00E-150	*Zea mays*	receptor-like kinase protein THICK TASSEL DWARF1
GR09G44800	HOM03D000009	ORTHO03D000261	479	3.00E-150	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
TC0007G01810	HOM03D000009	ORTHO03D000261	478	8.00E-150	*Theobroma cacao*	-
GM07G18590	HOM03D000009	ORTHO03D000261	475	1.00E-150	*Glycine max*	receptor-like kinase protein THICK TASSEL DWARF1
TC0010G16470	HOM03D000009	ORTHO03D012248	475	8.00E-150	*Theobroma cacao*	-
MT5G086810	HOM03D000009	ORTHO03D014880	475	4.00E-149	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR05G02940	HOM03D000009	ORTHO03D000261	475	9.00E-149	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
TC0007G01830	HOM03D000009	ORTHO03D088047	470	6.00E-147	*Theobroma cacao*	-
PT00G11850	HOM03D000009	ORTHO03D000261	469	5.00E-147	*Populus trichocarpa*	receptor-like kinase protein THICK TASSEL DWARF1

GR09G44810	HOM03D000009	ORTHO03D012248	469	2.00E-146	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR06G05630	HOM03D000009	ORTHO03D000261	468	8.00E-147	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR09G44790	HOM03D000009	ORTHO03D012248	468	1.00E-146	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
ZM03G04340	HOM03D000009	ORTHO03D000261	468	1.00E-145	*Zea mays*	receptor-like kinase protein THICK TASSEL DWARF1
BR09G09110	HOM03D000009	ORTHO03D035676	466	4.00E-146	*Brassica rapa*	receptor-like kinase protein FLORAL ORGAN NUMBER1
MD00G351510	HOM03D000009	ORTHO03D000261	465	1.00E-151	*Malus domestica*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR09G36770	HOM03D000009	ORTHO03D012248	463	3.00E-145	*Gossypium raimondii*	Brassinosteroid LRR receptor kinase
MT5G087320	HOM03D000009	ORTHO03D014880	463	1.00E-144	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
OS01G04070	HOM03D000009	ORTHO03D000261	463	1.00E-142	*Oryza sativa ssp. Japonica*	receptor-like kinase protein THICK TASSEL DWARF1
EG0002G05880	HOM03D000009	ORTHO03D007020	462	2.00E-144	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0002G12380	HOM03D000009	ORTHO03D007020	458	4.00E-144	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
EG0007G07160	HOM03D000009	ORTHO03D000261	457	6.00E-143	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0011G05350	HOM03D000009	ORTHO03D000261	457	2.00E-142	*Eucalyptus grandis*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0005G10240	HOM03D000009	ORTHO03D007020	456	2.00E-141	*Eucalyptus grandis*	receptor-like kinase protein THICK TASSEL DWARF1
VV12G08270	HOM03D000009	ORTHO03D014855	455	5.00E-144	*Vitis vinifera*	receptor-like kinase protein THICK TASSEL DWARF1
GR07G28110	HOM03D000009	ORTHO03D000261	455	4.00E-142	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR09G44820	HOM03D000009	ORTHO03D012248	455	4.00E-141	*Gossypium raimondii*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_004G24020	HOM03D000009	ORTHO03D000980	454	5.00E-143	*Prunus persica*	receptor-like kinase protein THICK TASSEL DWARF1
GR01G22340	HOM03D000009	ORTHO03D000261	454	1.00E-141	*Gossypium raimondii*	receptor-like kinase protein FLORAL ORGAN NUMBER1
RC29848G00730	HOM03D000009	ORTHO03D000261	453	1.00E-141	*Ricinus communis*	Receptor-like protein kinase BRI1-like 3
SL07G005150	HOM03D000009	ORTHO03D028741	453	2.00E-140	*Solanum lycopersicum*	repeat receptor protein kinase EXS
MT5G087070	HOM03D000009	ORTHO03D014880	453	5.00E-140	*Medicago truncatula*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0011G05380	HOM03D000009	ORTHO03D000261	452	9.00E-141	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
EG0011G05400	HOM03D000009	ORTHO03D000261	452	2.00E-140	*Eucalyptus grandis*	receptor-like kinase protein FLORAL ORGAN NUMBER1
OS04G28210	HOM03D000009	ORTHO03D000261	452	3.00E-140	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1
GR09G44860	HOM03D000009	ORTHO03D012248	452	4.00E-140	*Gossypium raimondii*	receptor-like kinase protein THICK TASSEL DWARF1
GM16G28480	HOM03D000009	ORTHO03D000980	451	4.00E-140	*Glycine max*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
TC0026G00080	HOM03D000009	ORTHO03D012248	451	5.00E-140	*Theobroma cacao*	-
OS01G06670	HOM03D000009	ORTHO03D000261	451	8.00E-140	*Oryza sativa ssp. Japonica*	receptor-like kinase protein FLORAL ORGAN NUMBER1

ST12G032880	HOM03D000009	ORTHO03D000980	450	9.00E-140	*Solanum tuberosum*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MT5G086530	HOM03D000009	ORTHO03D014880	449	4.00E-139	*Medicago truncatula*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MT5G086570	HOM03D000009	ORTHO03D014880	448	9.00E-138	*Medicago truncatula*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
ST10G018860	HOM03D000009	ORTHO03D000261	446	3.00E-140	*Solanum tuberosum*		receptor-like kinase protein THICK TASSEL DWARF1
EG0002G05900	HOM03D000009	ORTHO03D007020	446	2.00E-139	*Eucalyptus grandis*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
PPE_003G08160	HOM03D000009	ORTHO03D015309	446	6.00E-139	*Prunus persica*		receptor-like kinase protein FLORAL ORGAN NUMBER1
TC0007G11190	HOM03D000009	ORTHO03D028741	445	2.00E-137	*Theobroma cacao*		-
GR02G06050	HOM03D000009	ORTHO03D000980	445	4.00E-137	*Gossypium raimondii*		receptor-like kinase protein FLORAL ORGAN NUMBER1
MT5G087090	HOM03D000009	ORTHO03D014880	445	4.00E-137	*Medicago truncatula*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD15G019820	HOM03D000009	ORTHO03D000261	444	3.00E-144	*Malus domestica*		inactive leucine-rich repeat receptor-like protein kinase At3g28040
CRU_001G34610	HOM03D000009	ORTHO03D016978	441	1.00E-136	Capsella rubella		repeat receptor protein kinase EXS
AT1G45616	HOM03D000009	ORTHO03D016978	441	2.00E-136	*Arabidopsis thaliana*		receptor like protein 6
MT5G086550	HOM03D000009	ORTHO03D014880	441	3.00E-136	*Medicago truncatula*		inactive leucine-rich repeat receptor-like protein kinase At2g25790
MD00G485210	HOM03D000009	ORTHO03D000261	440	5.00E-143	*Malus domestica*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
GR11G15540	HOM03D000009	ORTHO03D000261	440	1.00E-135	*Gossypium raimondii*		receptor-like kinase protein FLORAL ORGAN NUMBER1
PT12G02760	HOM03D000009	ORTHO03D082383	439	3.00E-135	*Populus trichocarpa*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
BV5G12520	HOM03D000009	ORTHO03D007020	437	1.00E-135	*Beta vulgaris*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
EG0005G11420	HOM03D000009	ORTHO03D007020	436	2.00E-135	*Eucalyptus grandis*		receptor-like kinase protein THICK TASSEL DWARF1
MD09G008030	HOM03D000009	ORTHO03D058940	435	2.00E-140	*Malus domestica*		receptor-like kinase protein FLORAL ORGAN NUMBER1
GM03G07400	HOM03D000009	ORTHO03D000261	434	1.00E-135	*Glycine max*		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1
MD12G018150	HOM03D000009	ORTHO03D000261	423	4.00E-136	*Malus domestica*		receptor-like kinase protein FLORAL ORGAN NUMBER1

Supplementary Table 2.S4 Protein sequence analysis and classification for *Prupe.6G281400* reference and deletion-carrying allele.

Protein ID	AA start	AA end	Protein family library	Protein structural motif	Description	Signature match	Match evalue
*Prupe.6G281400*_REFERENCE	1	19	SignalP_EUK	SignalP-noTM			-
*Prupe.6G281400*_REFERENCE	1	19	Phobius	SIGNAL_PEPTIDE	Signal peptide region		-
*Prupe.6G281400*_REFERENCE	1	2	Phobius	SIGNAL_PEPTIDE_N_REGION	N-terminal region of a signal peptide.		-
*Prupe.6G281400*_REFERENCE	3	14	Phobius	SIGNAL_PEPTIDE_H_REGION	Hydrophobic region of a signal peptide.		-
*Prupe.6G281400*_REFERENCE	5	27	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.		-
*Prupe.6G281400*_REFERENCE	15	19	Phobius	SIGNAL_PEPTIDE_C_REGION	C-terminal region of a signal peptide.		-
*Prupe.6G281400*_REFERENCE	20	959	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.		-
*Prupe.6G281400*_REFERENCE	22	363	PANTHER	PTHR27004			1.6E-279
*Prupe.6G281400*_REFERENCE	22	363	PANTHER	PTHR27004:SF62			1.6E-279
*Prupe.6G281400*_REFERENCE	26	202	Gene3D	G3DSA:3.80.10.10	Leucine-rich repeat domain superfamily	IPR032675	2.2E-30
*Prupe.6G281400*_REFERENCE	33	73	Pfam	PF08263	Leucine-rich repeat-containing N-terminal, plant-type	IPR013210	9.0E-10
*Prupe.6G281400*_REFERENCE	80	390	SUPERFAMILY	SSF52047	Leucine-rich repeat domain superfamily		7.85E-45
*Prupe.6G281400*_REFERENCE	203	303	Gene3D	G3DSA:3.80.10.10	Leucine-rich repeat domain superfamily	IPR032675	6.3E-22
*Prupe.6G281400*_REFERENCE	232	255	SMART	SM00369	Leucine-rich repeat, typical subtype	IPR003591	37.0
*Prupe.6G281400*_REFERENCE	234	292	Pfam	PF13855	Leucine-rich repeat	IPR001611	9.5E-7
*Prupe.6G281400*_REFERENCE	256	280	SMART	SM00369	Leucine-rich repeat, typical subtype	IPR003591	220.0
*Prupe.6G281400*_REFERENCE	300	609	SUPERFAMILY	SSF52047			1.8E-47

*Prupe.6G281400*_REFERENCE	304	373	Gene3D	G3DSA:3.80.10.10	Leucine-rich repeat, typical subtype	IPR032675	2.6E-17
*Prupe.6G281400*_REFERENCE	356	440	PANTHER	PTHR27004			1.6E-279
*Prupe.6G281400*_REFERENCE	356	440	PANTHER	PTHR27004:SF62			1.6E-279
*Prupe.6G281400*_REFERENCE	374	708	Gene3D	G3DSA:3.80.10.10	Leucine-rich repeat domain superfamily	IPR032675	1.7E-70
*Prupe.6G281400*_REFERENCE	401	424	SMART	SM00369	Leucine-rich repeat, typical subtype	IPR003591	200.0
*Prupe.6G281400*_REFERENCE	436	509	PANTHER	PTHR27004			1.6E-279
*Prupe.6G281400*_REFERENCE	436	509	PANTHER	PTHR27004:SF62			1.6E-279
*Prupe.6G281400*_REFERENCE	471	495	SMART	SM00369		IPR003591	390.0
*Prupe.6G281400*_REFERENCE	516	1003	PANTHER	PTHR27004			1.6E-279
*Prupe.6G281400*_REFERENCE	516	1003	PANTHER	PTHR27004:SF62			1.6E-279
*Prupe.6G281400*_REFERENCE	570	593	SMART	SM00369	Leucine-rich repeat, typical subtype	IPR003591	190.0
*Prupe.6G281400*_REFERENCE	621	735	SUPERFAMILY	SSF52058			8.67E-41
*Prupe.6G281400*_REFERENCE	665	687	SMART	SM00369	Leucine-rich repeat, typical subtype	IPR003591	150.0
*Prupe.6G281400*_REFERENCE	666	685	Pfam	PF00560	Leucine-rich repeat	IPR001611	0.027
*Prupe.6G281400*_REFERENCE	667	680	PRINTS	PR00019	Leucine-rich repeat signature		2.5E-5
*Prupe.6G281400*_REFERENCE	815	935	SUPERFAMILY	SSF52058			8.67E-41
*Prupe.6G281400*_REFERENCE	845	869	SMART	SM00369	Leucine-rich repeat, typical subtype	IPR003591	42.0
*Prupe.6G281400*_REFERENCE	847	906	Pfam	PF13855	Leucine-rich repeat	IPR001611	1.7E-7
*Prupe.6G281400*_REFERENCE	869	882	PRINTS	PR00019	Leucine-rich repeat signature		2.5E-5
*Prupe.6G281400*_REFERENCE	893	926	SMART	SM00369	Leucine-rich repeat, typical subtype	IPR003591	66.0
*Prupe.6G281400*_REFERENCE	960	985	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.		-
*Prupe.6G281400*_REFERENCE	963	985	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.		-
*Prupe.6G281400*_REFERENCE	986	1023	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be		-

					outside the membrane, in the cytoplasm.	
M14_deletion-carrying allele	1	19	SignalP_EUK	SignalP-noTM		-
M14_deletion-carrying allele	1	19	Phobius	SIGNAL_PEPTIDE	Signal peptide region	-
M14_deletion-carrying allele	1	2	Phobius	SIGNAL_PEPTIDE_N_REGION	N-terminal region of a signal peptide.	-
M14_deletion-carrying allele	3	14	Phobius	SIGNAL_PEPTIDE_H_REGION	Hydrophobic region of a signal peptide.	-
M14_deletion-carrying allele	5	27	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	-
M14_deletion-carrying allele	15	19	Phobius	SIGNAL_PEPTIDE_C_REGION	C-terminal region of a signal peptide.	-
M14_deletion-carrying allele	20	409	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	-
M14_deletion-carrying allele	22	364	PANTHER	PTHR27004		3.1E-101
M14_deletion-carrying allele	22	364	PANTHER	PTHR27004:SF62		3.1E-101
M14_deletion-carrying allele	26	201	Gene3D	G3DSA:3.80.10.10	Leucine-rich repeat domain superfamily	3.9E-31	IPR032675
M14_deletion-carrying allele	33	73	Pfam	PF08263	Leucine-rich repeat-containing N-terminal, plant-type	2.9E-10	IPR013210
M14_deletion-carrying allele	64	408	SUPERFAMILY	SSF52058		1.43E-48
M14_deletion-carrying allele	202	302	Gene3D	G3DSA:3.80.10.10	Leucine-rich repeat domain superfamily	1.7E-22	IPR032675
M14_deletion-carrying allele	234	292	Pfam	PF13855	Leucine-rich repeat	2.9E-7	IPR001611
M14_deletion-carrying allele	303	409	Gene3D	G3DSA:3.80.10.10	Leucine-rich repeat domain superfamily	1.6E-24	IPR032675
M14_deletion-carrying allele	356	409	PANTHER	PTHR27004		3.1E-101
M14_deletion-carrying allele	356	409	PANTHER	PTHR27004:SF62		3.1E-101

Supplementary Sequence 3.S1 Sequence of the seven regions amplified with the primers designed to validate the loss of heterozygosis at the distal end of chromosome 6. SNP alleles were marked in bracket and bold. Yellow labelling indicated forward primers while reverse primers in red labelling.

1) Pp06:25256172..25256856 (LOH_Con2F/ LOH_Con2R)
ACTGCCAACATGTCCTCCTCCCCCTCCACCACCACCACCACCAAAATCTCCTTCACCGCCAAAACCAGCG
CCACCTCCTATGCCGCCGTCACCAAAACCAGCTCCACCGCCATAGTCAGCCCCACC**[G/T]**CCAAAGTCTG
TTCCAATGCCAATATTGGCTTCAACGTCGTTATTAATATGGGCTCCACCGCCTAAGCCTCCACCTCTACC
ACAGCCGGTACCACCATGGGGGACACTTGTACCGTTGTACGCCTTCCGAAAAACAAACTCAAAGCGATC
ATTACAGCAAAGGCCACAAACATAACTGTGACTATGATTGTCCACGTCTCCATCTTCATTTCTACTGGTA
TACGTTATATGTAACATGCCCATGGCAATGGCATCCCTTTTATAACCAATTACAAATATTAAAAGTGCCT
TTGGTGTCCAAAAAAAAGAAAAATTCAAGTTTACTTTTCCCTTTGACATGCGGAAGTTCAAGGTTTTGA
GTTTTCAAACATAATACAAATGCTTGGTAAGCATATATATGACCTCGTCTTTAAAAAATTTTAAAAATTA
AAAATAAAAATTGCTGATAAAATATGACTCAAATATTGATCTAGTCAAATAGAGGAGGCAAACGC

2) >Pp06:26227383..26228483 (LOH_Con3F/ LOH_Con3R)
ACAGAATTCAGCCGATGGAAGGGATTCCATTTCCTACACTCTCTTGGATGCCTCCAAGAAGGCATCATT
CTTAAATGCATTTGACAAAGCAGGATTTAAATCGTTAGATAAGGTCTTGGTTGCCTACAAACCTCGGAG
GGGGACGTTCGCAGTGTTTGAGGGTGAAATGACCACAGAAGAAGTAGAGAGGTTCATTGGCT**[C/T]**A
GTTCTTAATGGGGACATACGTTTTACCAAGACTCGGCAGAAACCCGTTCTGAAATGAGAGCTATGGTGA
CAAAGCTTGGGCCAGGTAGTTTCTGGCTTATTTGTACATAGCTTCACTATTTAAGTATAGAAAACATAAC
ACAAATTTAGACTAGAACTAAGGAGGATCTGTCAAGGAGCATCCGTCGTCATCTTGGAAAATTGGGCTT
GAACCCTACATAAGTGGTTCTTTTTGTAATGCCATAG**[T/C]**AGAAGTCTCTCTCATACATTT**[A/G]**GTTG
TATGACCCGTGTAATTTATTTGTTCTTTTATTCGGTGTAAAATTGATGACCAAAATATATATGAAAATGA
GAAGCTGATGAAGTTTGAAATTCAAGTTCTCGTTTTATCAAATACTGATTCGCTTCTTGCTGATTTTCATT
TATTTGC**[G/A]**ACCTGCA

3) >Pp06:26924276..26924785 (LOH_Con9F/ LOH_Con9R)
TTCTCGAATATGGGCTCCTGTATATGACAGGAATCTGGGCCCCCAAAGAAAAGAACCTGTTTGTCCATC
CTTGCAATTCAGCATAATGGGGCAAAAACTTTATGTCAGCCAAGAACAGGTAAGGAGTTTTTCTAGTAA
TCAATCTCTCCTCACTAAGAACAGGCAAGGACTTTTTTTAGCAACTTAATTGTCATGATAAAAACATA**[A/
T]**ATGCTTATAATTTAATTCCTCAAATGTAAAACACAATAGTCATGTCCTTGAGATGTCTAAATAGCCCTT
AAGAAACTGATGAAGGATACTTAACTAGATGACCCATTTGTTAATTTCGTGATCTCTCGTTCTGAGTTTG
ATTCGTGATTTAATGACATAGTGATAGTCCGTCCAAATTACAAATTGTCGTAGGTGATTTCATCTTAATC
CTTTTCAACAAAATTTTGCTACACACAGGTAACTGTTGGACGGAAGCCAGTGACAGGCGTACAGTTATG
TCTAGAAGGAAGCAAGCAGAATCGA

4) >Pp06:27488016..27488574 (LOH_Con10F/ LOH_Con10R)
GAGTCTTGGTCTGCATCTTGGTTTACACAAAAATAAAAAGAAAAAGAGAAGAAAAAAAAACATTATGG
ATA**[T/C]**AAAGTCTCTGAAACGCAAAAATCGAGATGAAAACAAAAAGGCCCAAACCCAAATTCCTAACA
ATCTTAGACTTGCAGAAGTTCTAAATGCTACAGTCGTCTTTAATTTTTGGGAAGGAAGAAAAGCCTCAT
GGCTATGAAGCTTGCCATGCCATCTGCATCGATGCCCTCAAACAACTTCTGGGCTTTTGGTCTACGAGG
GTCGTACTTCTTTTGAAATAAAAAGAGAACGCTCATCAACATGAAGCCTTTGCACCTGAGGCCCACAGT
TCTTTTTCTTTTTAAATTTGGTGGGCTCAGTTGTTAGCCACATGCTTCACTTTTTCCTTTTTTCTTCTTTTTC
TCTCTCTTTTTGATTCTTTTGCATATGAAAGGTTAATAAAAAGTTTACCAACAAAGTCTAATAACATGTGA
TTGGGACTTCTAATTTTCTTCAGATTCCACCATAACATAATAGTATGCACTATCCACTTCCTCATATGTCA
TGA

**5) >Pp06:28013592..28014146 (LOH_Con11F/ LOH_Con11R)**

AGATATTATACAAATTCATATGGTCATGTAAGGAATTAATGCATAACAT
TACAAAATTGAACCAATG**[A/T]**TAAGGATCCAATTTCCAAGTGATTTTATTAAAATATGAGAAGCCGTA
ATACATATGGGATGTCACAATAAGTAGGTAGATAAACAGGTCAGATATGCATTAAACACATAACAATTA
AACCAACTAAATGGATCTTATTTCCATGCATATGTGAAGCAACATAGATGAAGGCATGAAATTAAAGCA
AGCTTTATGAAAGCTGCTCCTCAAAGGCCAATTCCTCAATCATCATGTTCCCATGATTGAGGAGGGGAT
TGAAGGCTTCAATCCCAGAACACTTGTTGGGCAGTATTATAGGAGCAGCATTAAGCCCCTGAGAATCAA
TTGGTGATGAGTGCTGTGCAGGCAATGCGTATTTACGCGCAATCACGACTGAGTTAATGACTTCATCAG
TGGGATGAAACACAGAAAGAACCTCAAACCCTCTAAGATCATCACAAGGATCAAT<mark style="background:red">CACTGGATACAGA
AAGGCCC</mark>

6) >Pp06:28558467..28559118 (LOH_Con6F/ LOH_Con6R)

<mark style="background:yellow">TGACCCGTCCTTTTCTATGG</mark>ATTATATGATTCAAGAAACCTCGCCAGGGGCTTGTAGTTCAAGTAATTAAG
AATAATAATGTATCCTAGAT**[C/T]**TTAAGTCCAGTTCTTCTCTCTCCAATATTTCTTTGTATCAAAAAAAC
CTCAACCTATTTTTTTCTTGGTGAATACCTCCAACCCACTTGACCAAATAACCAAATATAGACATCAGATC
AAAAAAAAAAAAAAAAAAACAAATATAGACAAATTTTAGCAAGACTTGATTCTCTAATAACATCTGTATTT
GTTGAACAACTCCTGTCCTTTTCTGTTGTCTAGCTAGTGGCCCTTCGTCTACAATTTCCACCGCCCCGACA
GAATTTCAACTCCTCTGTTTTGTAGTTATTAGTTGTTCAAACCCTCTTTGTTTATATCTTTTGTGTCTCTATT
ATTTGCCATGTGTATGCATAGTATTGGATGATTTCAATTATGCTCATACTCAGGCTTAAAATGTAATTAG
CCACGACCAAAAAACACGTCCTTACCTTTCACGTAGTTGACAGAAACTCAATCTCTCTCGCTCACACTGT
ATGTGCCTAGGGTTTCTTTCAACAAGTGCCCAAGCAATTAACTACGACTAAAAACAACATCAGAATGGA
<mark style="background:red">GGATCTAATCGCCCTCAACA</mark>

7) >Pp06:29613605..29614163 (LOH_Con12F/ LOH_Con12R)

<mark style="background:yellow">TGCATCACAGCTCCAAAAAC</mark>CCAAAATTCCCATAAAAATCGAATTTTTCCCGGAAAAATAAAAATCCCCA
ACACCCTCATCTGAAATATTTTGTGCAACCTCAAAACCCCAATTTT**[G/A]**GCGGCGAAGTATCAAACCAA
CCCGAGAATTCAAATCCGGGTGTAAGAAGTCGGGCGCAGAGGCTGGTTTGCAACAACCGACACCAACC
CGTTTTCAACCGCAGTCTCCTCGTCAAGAAACAAGTCTTCGGTGACCCGAAAGGCGGCGTGGAGCCCA
ACCACCACCACCCCGACGATCAACGACACCAACACATTCAAACCCACGTGGGTAAACACCAGCGCTACG
ACAGTGACCAAGCTCAGCACGACCACCACCAGCCTGTCGTCGAAGCTCTGGTTGAACAAAACGACGGG
GCCAGTTCGGAAGAAGTAGAGGGCGAGCCAGGCGACGAGGACGATCAAGAAGACGATCATGGAGAC
TGGGTGCCATAGGAGGCTGAGGAAGACGATTAGGAGCACTGCCATGGCGTAGTTGACGCGGAAGTAG
CCGATGTTGTGCTTGATTCGAGCCATGGCGTCGGCGTAGT<mark style="background:red">TGTACGGAACGGAGAAGGAG</mark>
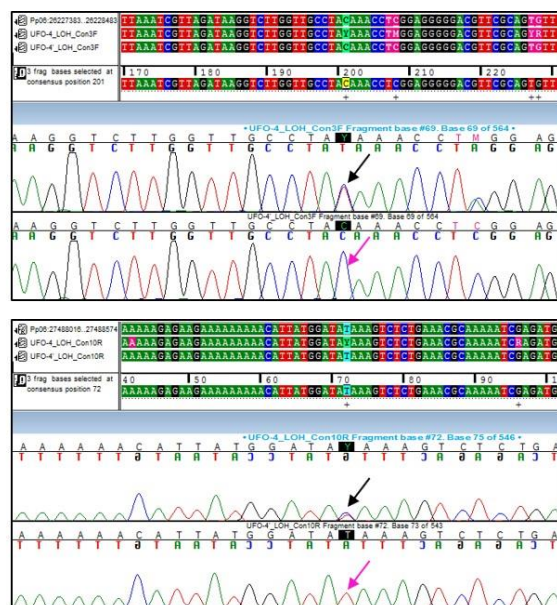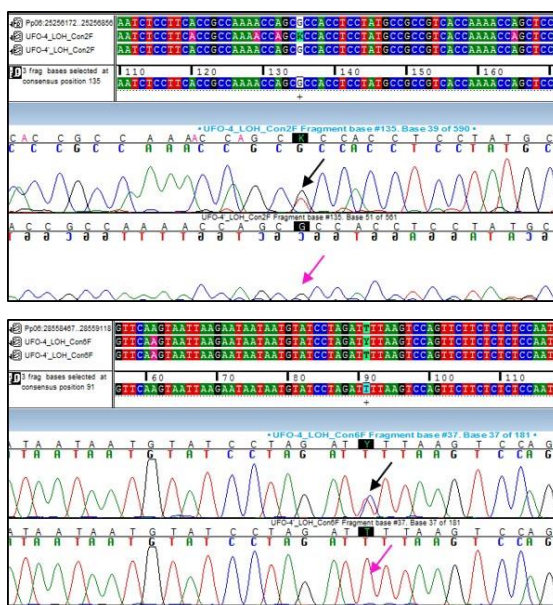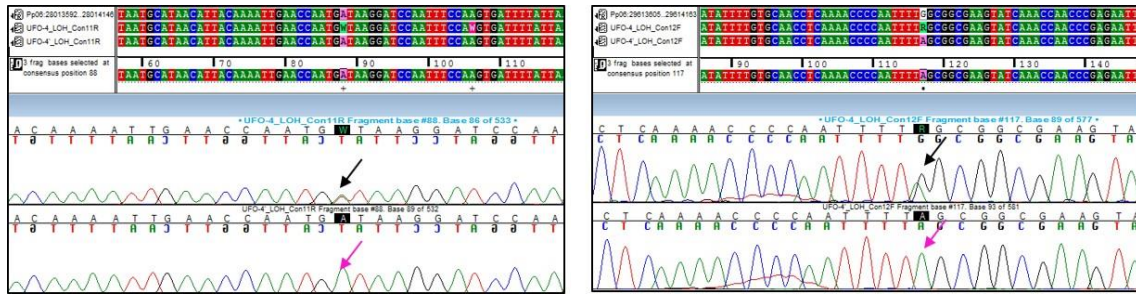
Supplementary Figure 3.S1 Overview of heterozygosis visualization based on Sanger analysis. SNP loci show heterozygosis in UFO-4 but homozygosis in UFO-4Mut. Those amplicons sequenced with forward primers are presented at the above in four shadowing colors by A, T, C, G, corresponded to below chromatograms. Black arrows indicate heterozygotes, and homozygotes in pink arrows.

Supplementary Sequence 4.S1 Sequence of plasmid pSNPPV5'BD-GFP which was digested with BamHI and XbaI for new recombinants construction. The restriction enzyme cutting sites and recombinant sites were marked and illustrated. All the fragments were used for hybridization with tailed primers and Gibson assembly.

Vector (22125 bp)
(XbaI restriction site)
CTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGT
TATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATG
AGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCCAAAGACAAAAGGG
CGACATTCAACCGATTGAGGGAGGGAAGGTAAATATTGACGGAAATTATTCATTAAAGGTGAATTATC
ACCGTCACCGACTTGAGCCATTTGGGAATTAGAGCCAGCAAAATCACCAGTAGCACCATTACCATTAGC
AAGGCCGGAAACGTCACCAATGAAACCATCGATAGCAGCACCGTAATCAGTAGCGACAGAATCAAGTT
TGCCTTTAGCGTCAGACTGTAGCGCGTTTTCATCGGCATTTTCGGTCATAGCCCCCCTTATTAGCGTTTGC
CATCTTTTCATAATCAAAATCACCGGAACCAGAGCCACCACCGGAACCGCCTCCCTCAGAGCCGCCACC
CTCAGAACCGCCACCCTCAGAGCCACCACCCTCAGAGCCGCCACCAGAACCACCACCAGAGCCGCCGCC
AGCATTGACAGGAGGCCCGATCTAGTAACATAGATGACACCGCGCGCGATAATTTATCCTAGTTTGCGC
GCTATATTTTGTTTTCTATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATCTCATA
AATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAATTCAACAGAAATTATATGATAATCA
TCGCAAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTGCCAAATGTTTGAACGATCGGGGATCA
TCCGGGTCTGTGGCGGGAACTCCACGAAAATATCCGAACGCAGCAAGATATCGCGGTGCATCTCGGTC
TTGCCTGGGCAGTCGCCGCCGACGCCGTTGATGTGGACGCCGGGCCCGATCATATTGTCGCTCAGGAT
CGTGGCGTTGTGCTTGTCGGCCGTTGCTGTCGTAATGATATCGGCACCTTCGACCGCCTGTTCCGCAGA
GATCCCGTGGGCGAAGAACTCCAGCATGAGATCCCCGCGCTGGAGGATCATCCAGCCGGCGTCCCGGA
AAACGATTCCGAAGCCCAACCTTTCATAGAAGGCGGCGGTGGAATCGAAATCTCGTGATGGCAGGTTG
GGCGTCGCTTGGTCGGTCATTTCGAACCCCAGAGTCCCGCTCAGAAGAACTCGTCAAGAAGGCGATAG
AAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTCAGCCCATTCGC
CGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACACCCAGCC
GGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGCAGGCATCGCCAT
GGGTCACGACGAGATCATCGCCGTCGGGCATGCGCGCCTTGAGCCTGGCGAACAGTTCGGCTGGCGC
GAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCG
CTCGATGCGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGCA
TTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGC
ACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACAACGTCGAGCACAGCTGCGCAAGGAACG
CCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCCTGCAGTTCATTCAGGGCACCGGACAGGTC
GGTCTTGACAAAAAGAACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGCCG
ATTGTCTGTTGTGCCCAGTCATAGCCGAATAGCCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAAT
CCATCTTGTTCAATCATGCGAAACGATCCAGATCCGGTGCAGATTATTTGGATTGAGAGTGAATATGAG
ACTCTAATTGGATACCGAGGGGAATTTATGGAACGTCAGTGGAGCATTTTTGACAAGAAATATTTGCTA
GCTGATAGTGACCTTAGGCGACTTTTGAACGCGCAATAATGGTTTCTGACGTATGTGCTTAGCTCATTA
AACTCCAGAAACCCGCGGCTGAGTGGCTCCTTCAACGTTGCGGTTCTGTCAGTTCCAAACGTAAAACGG
CTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCTTAATTCTCCGCTCATGATCAGATTGTCGT
TTCCCGCCTTCAGTTTAAACTATCAGTGTTTGACAGGATATATTGGCGGGTAAACCTAAGAGAAAAGAG
CGTTTATTAGAATAATCGGATATTTAAAAGGGCGTGAAAAGGTTTATCCGTTCGTCCATTTGTATGTGCA

TGCCAACCACAGGGTTCCCCAGATCTGGCGCCGGCCAGCGAGACGAGCAAGATTGGCCGCCGCCCGAA
ACGATCCGACAGCGCGCCCAGCACAGGTGCGCAGGCAAATTGCACCAACGCATACAGCGCCAGCAGA
ATGCCATAGTGGGCGGTGACGTCGTTCGAGTGAACCAGATCGCGCAGGAGGCCCGGCAGCACCGGCA
TAATCAGGCCGATGCCGACAGCGTCGAGCGCGACAGTGCTCAGAATTACGATCAGGGGTATGTTGGGT
TTCACGTCTGGCCTCCGGACCAGCCTCCGCTGGTCCGATTGAACGCGCGGATTCTTTATCACTGATAAGT
TGGTGGACATATTATGTTTATCAGTGATAAAGTGTCAAGCATGACAAAGTTGCAGCCGAATACAGTGAT
CCGTGCCGCCCTGGACCTGTTGAACGAGGTCGGCGTAGACGGTCTGACGACACGCAAACTGGCGGAA
CGGTTGGGGGTTCAGCAGCCGGCGCTTTACTGGCACTTCAGGAACAAGCGGGCGCTGCTCGACGCACT
GGCCGAAGCCATGCTGGCGGAGAATCATACGCATTCGGTGCCGAGAGCCGACGACGACTGGCGCTCA
TTTCTGATCGGGAATGCCCGCAGCTTCAGGCAGGCGCTGCTCGCCTACCGCGATGGCGCGCGCATCCAT
GCCGGCACGCGACCGGGCGCACCGCAGATGGAAACGGCCGACGCGCAGCTTCGCTTCCTCTGCGAGG
CGGGTTTTTCGGCCGGGGACGCCGTCAATGCGCTGATGACAATCAGCTACTTCACTGTTGGGGCCGTG
CTTGAGGAGCAGGCCGGCGACAGCGATGCCGGCGAGCGCGGCGGCACCGTTGAACAGGCTCCGCTCT
CGCCGCTGTTGCGGGCCGCGATAGACGCCTTCGACGAAGCCGGTCCGGACGCAGCGTTCGAGCAGGG
ACTCGCGGTGATTGTCGATGGATTGGCGAAAAGGAGGCTCGTTGTCAGGAACGTTGAAGGACCGAGA
AAGGGTGACGATTGATCAGGACCGCTGCCGGAGCGCAACCCACTCACTACAGCAGAGCCATGTAGACA
ACATCCCCTCCCCCTTTCCACCGCGTCAGACGCCCGTAGCAGCCCGCTACGGGCTTTTTCATGCCCTGCC
CTAGCGTCCAAGCCTCACGGCCGCGCTCGGCCTCTCTGGCGGCCTTCTGGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATAC
GGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAG
GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAA
TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA
GCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCC
TTTTTCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTC
AGCCGGGCAGGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGC
ACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTAACAGATGAG
GGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCC
TTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTGTCGGCCTACCT
GCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCC
CGCATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCAC
GGCGCGGTTCGGTGATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTT
GGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAACGGCC
GGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGCGACTTCGCGGAGCT
GGTGAAGTACATCACCGACGAGCAAGGCAAGACCGAGCGCCTTTGCGACGCTCACCGGGCTGGTTGCC
CTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGC
GAGACACCGCGGCCGCCGGCGTTGTGGATACCTCGCGGAAAACTTGGCCCTCACTGACAGATGAGGG
GCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGACAGATGAGGGGCAGGCTCGAT
TTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTTACGCGAGTTTC
CCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGGGCGCGACTAC
TGACAGATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTA
TTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGCCCGTTTTTCGGC
CACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAACCAGGGCTGCGC
CCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCCTCCCGGCCCGCTAAC
GCGGGCCTCCCATCCCCCCAGGGGCTGCGCCCCTCGGCCGCGAACGGCCTCACCCCAAAAATGGCAGC

GCTGGCAGTCCTTGCCATTGCCGGGATCGGGGCAGTAACGGGATGGGCGATCAGCCCGAGCGCGACG
CCCGGAAGCATTGACGTGCCGCAGGTGCTGGCATCGACATTCAGCGACCAGGTGCCGGGCAGTGAGG
GCGGCGGCCTGGGTGGCGGCCTGCCCTTCACTTCGGCCGTCGGGGCATTCACGGACTTCATGGCGGGG
CCGGCAATTTTTACCTTGGGCATTCTTGGCATAGTGGTCGCGGGTGCCGTGCTCGTGTTCGGGGGGTGCG
ATAAACCCAGCGAACCATTTGAGGTGATAGGTAAGATTATACCGAGGTATGAAAACGAGAATTGGACC
TTTACAGAATTACTCTATGAAGCGCCATATTTAAAAAGCTACCAAGACGAAGAGGATGAAGAGGATGA
GGAGGCAGATTGCCTTGAATATATTGACAATACTGATAAGATAATATATCTTTTATATAGAAGATATCG
CCGTATGTAAGGATTTCAGGGGGGCAAGGCATAGGCAGCGCGCTTATCAATATATCTATAGAATGGGCA
AAGCATAAAAACTTGCATGGACTAATGCTTGAAACCCAGGACAATAACCTTATAGCTTGTAAATTCTATC
ATAATTGGGTAATGACTCCAACTTATTGATAGTGTTTTATGTTCAGATAATGCCCGATGACTTTGTCATG
CAGCTCCACCGATTTTGAGAACGACAGCGACTTCCGTCCCAGCCGTGCCAGGTGCTGCCTCAGATTCAG
GTTATGCCGCTCAATTCGCTGCGTATATCGCTTGCTGATTACGTGCAGCTTTCCCTTCAGGCGGGATTCA
TACAGCGGCCAGCCATCCGTCATCCATATCACCACGTCAAAGGGTGACAGCAGGCTCATAAGACGCCCC
AGCGTCGCCATAGTGCGTTCACCGAATACGTGCGCAACAACCGTCTTCCGGAGACTGTCATACGCGTAA
AACAGCCAGCGCTGGCGCGATTTAGCCCCGACATAGCCCCACTGTTCGTCCATTTCCGCGCAGACGATG
ACGTCACTGCCCGGCTGTATGCGCGAGGTTACCGACTGCGGCCTGAGTTTTTTAAGTGACGTAAAATCG
TGTTGAGGCCAACGCCCATAATGCGGGCTGTTGCCCGGCATCCAACGCCATTCATGGCCATATCAATGA
TTTTCTGGTGCGTACCGGGTTGAGAAGCGGTGTAAGTGAACTGCAGTTGCCATGTTTTACGGCAGTGA
GAGCAGAGATAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACA
GGAGGGACAGCTGATAGACACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGT
AAGTTGGCAGCATCACCCATAATTGTGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGCCAACTT
TGAAAACAACTTTGAAAAAGCTGTTTTCTGGTATTTAAGGTTTTAGAATGCAAGGAACAGTGAATTGGA
GTTCGTCTTGTTATAATTAGCTTCTTGGGGTATCTTTAAATACTGTAGAAAAGAGGAAGGAAATAATAA
ATGGCTAAAATGAGAATATCACCGGAATTGAAAAAACTGATCGAAAAATACCGCTGCGTAAAAGATAC
GGAAGGAATGTCTCCTGCTAAGGTATATAAGCTGGTGGGAGAAAATGAAAACCTATATTTAAAAATGA
CGGACAGCCGGTATAAAGGGACCACCTATGATGTGGAACGGGAAAAGGACATGATGCTATGGCTGGA
AGGAAAGCTGCCTGTTCCAAAGGTCCTGCACTTTGAACGGCATGATGGCTGGAGCAATCTGCTCATGA
GTGAGGCCGATGGCGTCCTTTGCTCGGAAGAGTATGAAGATGAACAAAGCCCTGAAAAGATTATCGAG
CTGTATGCGGAGTGCATCAGGCTCTTTCACTCCATCGACATATCGGATTGTCCCTATACGAATAGCTTAG
ACAGCCGCTTAGCCGAATTGGATTACTTACTGAATAACGATCTGGCCGATGTGGATTGCGAAAACTGG
GAAGAAGACACTCCATTTAAAGATCCGCGCGAGCTGTATGATTTTTTAAAGACGGAAAAGCCCGAAGA
GGAACTTGTCTTTTCCCACGGCGACCTGGGAGACAGCAACATCTTTGTGAAAGATGGCAAAGTAAGTG
GCTTTATTGATCTTGGGAGAAGCGGCAGGGCGGACAAGTGGTATGACATTGCCTTCTGCGTCCGGTCG
ATCAGGGAGGATATCGGGGAAGAACAGTATGTCGAGCTATTTTTTGACTTACTGGGGATCAAGCCTGA
TTGGGAGAAAATAAAATATTATATTTTACTGGATGAATTGTTTTAGTACCTAGATGTGGCGCAACGATG
CCGGCGACAAGCAGGAGCGCACCGACTTCTTCCGCATCAAGTGTTTTGGCTCTCAGGCCGAGGCCCAC
GGCAAGTATTTGGGCAAGGGGTCGCTGGTATTCGTGCAGGGCAAGATTCGGAATACCAAGTACGAGA
AGGACGGCCAGACGGTCTACGGGACCGACTTCATTGCCGATAAGGTGGATTATCTGGACACCAAGGCA
CCAGGCGGGTCAAATCAGGAATAAGGGCACATTGCCCCGGCGTGAGTCGGGGCAATCCCGCAAGGAG
GGTGAATGAATCGGACGTTTGACCGGAAGGCATACAGGCAAGAACTGATCGACGCGGGGTTTTCCGC
CGAGGATGCCGAAACCATCGCAAGCCGCACCGTCATGCGTGCGCCCCGCGAAACCTTCCAGTCCGTCG
GCTCGATGGTCCAGCAAGCTACGGCCAAGATCGAGCGCGACAGCGTGCAACTGGCTCCCCCTGCCCTG
CCCGCGCCATCGGCCGCCGTGGAGCGTTCGCGTCGTCTCGAACAGGAGGCGGCAGGTTTGGCGAAGT
CGATGACCATCGACACGCGAGGAACTATGACGACCAAGAAGCGAAAAACCGCCGGCGAGGACCTGGC
AAAACAGGTCAGCGAGGCCAAGCAGGCCGCGTTGCTGAAACACACGAAGCAGCAGATCAAGGAAATG

CAGCTTTCCTTGTTCGATATTGCGCCGTGGCCGGACACGATGCGAGCGATGCCAAACGACACGGCCCG
CTCTGCCCTGTTCACCACGCGCAACAAGAAAATCCCGCGCGAGGCGCTGCAAAACAAGGTCATTTTCCA
CGTCAACAAGGACGTGAAGATCACCTACACCGGCGTCGAGCTGCGGGCCGACGATGACGAACTGGTG
TGGCAGCAGGTGTTGGAGTACGCGAAGCGCACCCCTATCGGCGAGCCGATCACCTTCACGTTCTACGA
GCTTTGCCAGGACCTGGGCTGGTCGATCAATGGCCGGTATTACACGAAGGCCGAGGAATGCCTGTCGC
GCCTACAGGCGACGGCGATGGGCTTCACGTCCGACCGCGTTGGGCACCTGGAATCGGTGTCGCTGCTG
CACCGCTTCCGCGTCCTGGACCGTGGCAAGAAAACGTCCCGTTGCCAGGTCCTGATCGACGAGGAAAT
CGTCGTGCTGTTTGCTGGCGACCACTACACGAAATTCATATGGGAGAAGTACCGCAAGCTGTCGCCGA
CGGCCCGACGGATGTTCGACTATTTCAGCTCGCACCGGGAGCCGTACCCGCTCAAGCTGGAAACCTTCC
GCCTCATGTGCGGATCGGATTCCACCCGCGTGAAGAAGTGGCGCGAGCAGGTCGGCGAAGCCTGCGA
AGAGTTGCGAGGCAGCGGCCTGGTGGAACACGCCTGGGTCAATGATGACCTGGTGCATTGCAAACGC
TAGGGCCTTGTGGGGTCAGTTCCGGCTGGGGGTTCAGCAGCCAGCGCTTTACTGGCATTTCAGGAACA
AGCGGGCACTGCTCGACGCACTTGCTTCGCTCAGTATCGCTCGGGACGCACGGCGCGCTCTACGAACT
GCCGATAAACAGAGGATTAAAATTGACAATTGTGATTAAGGCTCAGATTCGACGGCTTGGAGCGGCCG
ACGTGCAGGATTTCCGCGAGATCCGATTGTCGGCCCTGAAGAAAGCTCCAGAGATGTTCGGGTCCGTT
TACGAGCACGAGGAGAAAAAGCCCATGGAGGCGTTCGCTGAACGGTTGCGAGATGCCGTGGCATTCG
GCGCCTACATCGACGGCGAGATCATTGGGCTGTCGGTCTTCAAACAGGAGGACGGCCCCAAGGACGCT
CACAAGGCGCATCTGTCCGGCGTTTTCGTGGAGCCCGAACAGCGAGGCCGAGGGGTCGCCGGTATGCT
GCTGCGGGCGTTGCCGGCGGGTTTATTGCTCGTGATGATCGTCCGACAGATTCCAACGGGAATCTGGT
GGATGCGCATCTTCATCCTCGGCGCACTTAATATTTCGCTATTCTGGAGCTTGTTGTTTATTTCGGTCTAC
CGCCTGCCGGGCGGGGTCGCGGCGACGGTAGGCGCTGTGCAGCCGCTGATGGTCGTGTTCATCTCTGC
CGCTCTGCTAGGTAGCCCGATACGATTGATGGCGGTCCTGGGGGGCTATTTGCGGAACTGCGGGCGTGG
CGCTGTTGGTGTTGACACCAAACGCAGCGCTAGATCCTGTCGGCGTCGCAGCGGGCCTGGCGGGGGGC
GGTTTCCATGGCGTTCGGAACCGTGCTGACCCGCAAGTGGCAACCTCCCGTGCCTCTGCTCACCTTTACC
GCCTGGCAACTGGCGGCCGGAGGACTTCTGCTCGTTCCAGTAGCTTTAGTGTTTGATCCGCCAATCCCG
ATGCCTACAGGAACCAATGTTCTCGGCCTGGCGTGGCTCGGCCTGATCGGAGCGGGTTTAACCTACTTC
CTTTGGTTCCGGGGGGATCTCGCGACTCGAACCTACAGTTGTTTCCTTACTGGGCTTTCTCAGCCCCAGAT
CTGGGGTCGATCAGCCGGGGATGCATCAGGCCGACAGTCGGAACTTCGGGTCCCCGACCTGTACCATT
CGGTGAGCAATGGATAGGGGAGTTGATATCGTCAACGTTCACTTCTAAAGAAATAGCGCCACTCAGCT
TCCTCAGCGGCTTTATCCAGCGATTTCCTATTATGTCGGCATAGTTCTCAAGATCGACAGCCTGTCACGG
TTAAGCGAGAAATGAATAAGAAGGCTGATAATTCGGATCTCTGCGAGGGAGATGATATTTGATCACAG
GCAGCAACGCTCTGTCATCGTTACAATCAACATGCTACCCTCCGCGAGATCATCCGTGTTTCAAACCCGG
CAGCTTAGTTGCCGTTCTTCCGAATAGCATCGGTAACATGAGCAAAGTCTGCCGCCTTACAACGGCTCT
CCCGCTGACGCCGTCCCGGACTGATGGGCTGCCTGTATCGAGTGGTGATTTTGTGCCGAGCTGCCGGTC
GGGGAGCTGTTGGCTGGCTGGTGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACTTA
ATAACACATTGCGGACGTTTTTAATGTACTGGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGC
TGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAG
GCGAAAATCCTGTTTGATGGTGGTTCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAG
ATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAA
GGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGG
GTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGA
AAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCCATTCAGGCTGC
GCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATG
TGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCA
GTGAATTCGAGCTCGGTACCCCTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCGGAATTCG

CCCGGGATCTCCTTTGCCCCAGAGATCACAATGGACGACTTCCTATATCTCTACGATCTAGTCAGGAAGT
TCGACGGAGAAGGTGACGATACCATGTTCACCACTGATAATGAGAAGATTAGCCTTTTCAATTTCAGAA
AGAATCCTAACCCACAGATGGTTAGAGACGCTTACGCAGCAGGTCTCATCAAGACGATCTACCCGAGC
AATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAAGATTCAGGACTAAC
TGCATCAAGAACACAGAGAAAGATATATTTCTCAAGATCAGAAGTACTCACCAGTCACAGAAAAGCATC
TTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCC
AACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCAT
GTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCAC
GATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCG
GCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGG
CTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTG
GGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGA
ACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTA
CTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTG
ATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGAT
CAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTA
CCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGA
GCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA
CCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTA
CCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTG
CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAA
GCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAG
AGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCT
GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT
GTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCG
ATTCATTAATGCAGCTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCGGAATTCGCCCGGGA
TCTCCTTTGCCCCAGAGATCACAATGGACGACTTCCTATATCTCTACGATCTAGTCAGGAAGTTCGACGG
AGAAGGTGACGATACCATGTTCACCACTGATAATGAGAAGATTAGCCTTTTCAATTTCAGAAAGAATCC
TAACCCACAGATGGTTAGAGACGCTTACGCAGCAGGTCTCATCAAGACGATCTACCCGAGCAATAATCT
CCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAAGATTCAGGACTAACTGCATCA
AGAACACAGAGAAAGATATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAAGGCTTGC
TTCACAAACCAAGGCAAGTAATAGAGATTGGAGTCTCTAAAAAGGTAGTTCCCACTGAATCAAAGGCC
ATGGAGTCAAAGATTCAAATAGAGGACCTAACAGAACTCGCCGTAAAGACTGGCGAACAGTTCATACA
GAGTCTCTTACGACTCAATGACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGACACGCTTGTCTA
CCTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGAATTGAGACTTTTCAACAAAGGGTAA
TATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGG
AAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGAC
AGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTC
TTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACGCTTGTCTACCTCCAAAAATATCA
AAGATACAGTCTCAGAAGACCAAAGGGAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCC
TCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACA
AATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGAT
GGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGA

TTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCT
ATATAAGGAAGTTCATTTCATTTGGAGAGGAAAATATAAAAACTCAACACAACATACAAAATTTTATGC
GATCAAATCAATCTCAAGCTATCAAAATTTTTCAAATCTCACTTGAAAGATCAAAAATCAACAAAGAAAA
TCCCTTAATTTCTCTACCAAATTTACTGCAAGTCAAGATGTCAACCATTGTATTTGGCTCATTCACTTGCC
ACCTCGATGCAGCTATCCACCAGGATAATGCAGACAGATTGGCAAAGGCCTGGACCCGTCCAGAGAAC
CGCCAAGTCAGTAACGTGCATCTACTGTGCCGAAGAGCGGCAAAAAGTCTCATAAACACATATGAGAG
TGCAACAGCTAGTGCTTGGAAAGGGCTGGAAGAGAAGTTGCAACCTATGTTTGCTAAGCGTGAGTTTA
GCAAAACAGTCACAAAGAGAAAAGGGCTACGGTGCTTCAAAGAAAGCTCTGAGAAGTTTATTGAAAA
GAAGCTCAGGAAACAGTATCAAGAGGAGCGTGAGAGATTCCAATTTCTCAACGGTCCGGATGCAATAG
TCAACCAAATCAGTGTTGACAAATGTGAAGCTTCAGTATGGGTGCCATTCCCTCATATTATTGAGAAAC
CTAGCTTTGCAACACCATCAATGAAAAAGAAGGTGGTGTTTACTAAGGTTAGGATGTCCGAGGCATCAC
TACAGCTTTTTATGAGGAGGGTTGCTGCAAACGCCAAGGCAAATGGTCAAAAAGTTGAGATCATAGGG
CGTAAGCGTGTAGTCGGTAACTACACAACGAAAAGTCGCCTGACATACTTTCGCACACATGTTCGGCAC
TTGGATGGTTCAAAACCACGCTATGATCTTGTGTTGGACGAGGCAACCAAGAAGATTCTGCAACTGTTT
GCAAACACAAGCGGTTTCCGCCATGTCCACAAGAAAGGGGAGGTAACACCAGGAATGAGCGGATTTG
TGGTAAATCCCATAAATCTATCGGACCCAATGCAAGTGTATGACACGGATCTTTTTATAGTTCGTGGAA
AACACAACTCTATTCTTGTTGACTCACGGTGTAAGGTTTCTAAAAAACAGAGCAATGAGATAATCCACT
ACTCTGACCCAGGCAAACAATTTTGGGATGGTTTCACCAATTCATTTATGCAGTGCAAGCTACGCGAAA
CTGATCATCAGTGCACATCTGACCTGGACGTGAAGGAGTGTGGTTATGTCGCAGCACTTGTGTGCCAA
GCGATAATCCCTTGCGGAAAAATCACATGTCTGCAATGTGCTCAAAAGTATTCTTACATGTCACAACAG
GAAATACGTGATAGATTTTCAACAGTAATTGAGCAGCATGAGAAAACAGTGATGGATAACTATCCACA
ATTTTCACATGTTCTCGCTTTTCTAAAGAGATATCGTGAACTAATGCGCGTGGAAAATCAGAATTATGAA
GCTTTCAAGGATATCACGCACATGATAGGCGAGCGTAAAGAAGCACCTTTTTCTCATCTCAACAAAATC
AATGAATTAATCATTAAGGGTGGTATGATGAGCGCACAAGACTACATAGAAGCCTCGGATCATCTGCG
CGAACTAGCGCGATATCAGAAGAATCGCACGGAGAACATTAGGAGCGGATCTATAAAGGCTTTCAGGA
ATAAAATCTCATCAAAAGCACATGTTAATATGCAGCTTATGTGTGACAATCAACTTGATACTAATGGCAA
TTTCGTGTGGGGACAGAGAGAGTATCATGCTAAACGCTTCTTTAGGAATTACTTCGATGTGATCGATGT
TAGCGAGGGCTACAGACGTCATATTGTTCGTGAAAATCCTAGAGGTATCCGCAAATTGGCCATTGGCA
ACCTTGTTATGTCAACAAATCTGGCAGCACTACGTAAGCAGCTTTTGGGTGAAGAGTGCATTCATTTTG
AGGTCTCAAAGGAATGCACTAGCAGGCGAGGGGAAAACTTTGTATACCAATGTTGCTGTGTCACACAT
GAAGACGGTACACCACTGGAGTCTGAAATAATAAGTCCAACAAAGAATCATTTAGTTGTTGGTAACTCA
GGTGATTCGAAGTATGTGGATTTGCCCACAGCAAAAGGAGGTGCAATGTTCATAGCAAAGGCAGGTTA
TTGTTACATTAACATTTTCCTTGCTATGCTGATCAACATAAATGAAGATGAAGCAAAAAGTTTCACAAAG
ACAGTGCGTGACACTCTTGTACCTAAGCTTGGAACATGGCCATCGATGATGGACTTAGCTACAGCTTGC
CACTTTCTCGCAGTTCTCTACCCAGAAACTCGGAATGCTGAGCTTCCACGAATACTCGTTGATCATGAAG
CAAAAATCTTTCATGTAGTTGACTCATTCGGATCACTGTCAACTGGAATGCATGTTTTGAAAGCGAACAC
AATCAACCAGCTTATTAGCTTTGCTAGTGATACATTGGATTCAAATATGAAAACATACCTGGTTGGAGG
TCTTGAAGTGGATAAGTGTGACGAATTCAAAAATGTCAAGCTCTTGATCAGAAGCATTTACAAGCCACA
AATCATGGAGCAGGTGCTTAAGGAAGAACCATATTTATTGCTCATGAGCGTTTTGTCACCTGGCGTCTT
GATGGCGCTGTTCAATAGTGGTTCATTGGAGAAAGCCACACAATATTGGATCACACGATCTCATAGCTT
GGCAGCGATCACATCAATGTTATCAGCACTTGCAGCCAAAGTTTCACTCGCAAGTACACTGAATGCACA
GATGAGTGTCATTGACGAACATGCAGCAGTTCTATGTGATAGTGTTTTTGATGGAACGAAGCCATACGC
ATCCTACATGATGGCAGTGAAAACTTTGGAAAGGATGAAAGCACGGACGGAATCCGATCACACTTTAA
ATGACTTAGGATTCTCGGTACTCAGACAGGCAACACCCCACTTGGTTGAAAAAAGTTATCTGCAGGTAA
GTTTCTGCTTCTACCTTTGATATATATATAATAATTATCATTAATTAGTAGTAATATAATATTTCAAATATT

TTTTTCAAAATAAAAGAATGTAGTATATAGCAATTGCTTTTCTGTAGTTTATAAGTGTGTATATTTTAATT
TATAACTTTTCTAATATATGACCAAAATTTGTTGATATGCAGGAATTGGAGCAAGCTTGGAAAGAGTTA
AGCTGGTCGGAAAAATTCTCTGCAATCTTGGAATCGCAGCGGTGGCGAAAACATATACCAAAACCTTTC
ATCCCAAAAGACGGCGCAGATTTAGGAGGCAGGTACGACATCTCCGTTCGGTCATTACTTGGCAACCA
GTACAAACGCCTGAGAGACGTAGTCCGACGGAAAAGAGACGATGTGGTTTGCTATACACACCAGTCGA
TGGGGAAGCTGTTTTGCAAAGCCATCGGAATTTCCACAAGTTTTCTTCCAAGCACTCTTAAAATGCTTGA
CATGCTCATCGTGTTCGGTCTCTTGCTTTCAATAGGAGCCACATGCAACTCAATGATCAATGAGCATAAA
CATCTTAAGCAACTTGCCGCCGATCGGGAAGATAAGAAAAGATTCAAAAGATTGCAAGTCTTATACACG
AGGTTATCAGAGAAAGTTGGTTGCACACCAACAGCAGATGAATTCCTGGAGTATGTGGGAGGTGAAAA
CCCTGATTTACTGAAACATGCAGAGGATCTAATTGGGGATGGTCAAGTTGTTGTTCATCAAAGCAAGAG
AGACTCACAAGCAAATTTGGAACGGGTTGTAGCATTTGTGGCTCTTGTTATGATGCTGTTTGACTCGGA
GCGAAGTGACGGCGTGTACAAGATTCTCAATAAACTCAAAGGCATTATGGGGAGTGTCGACCAGGCTG
TTCATCATCAGAGCTTGGACGATATAGAAGATATACTGGATGAGAAGAAGCTCACAGTCGATTTTGTAC
TGCAAAGTAACGAAGTTGCACCAACTGTCCCATTTGACTCAACTTTTGAGAAATGGTGGACGAATCAAC
TTGAGACAGGAAATGTGATTCCACACTACAGGACTGAAGGACATTTCCTTGAATTCACACGAGAAAATG
CAGCGCACATTGCGAATGAAGTCATGCATGGCTCACATCAAGATATCCTAATTCGTGGAGCAGTTGGAT
CGGGCAAATCAACTGGATTGCCATTCCACTTAAGCAAGAAGGGCCACGTCCTGCTAATTGAACCCACCA
GGCCGCTAGCTGAGAATGTGTGCAAGCAGTTGCGAGGTCAACCATTCAATGTCAATCCTACACTGCGC
ATGCGTGGGATGAGCACCTTTGGATCAACTCCAATTACTGTGATGACAAGCGGTTACGCACTGCACTTC
TTGGCAAACAATCCAACTTATTTGGACAACTATAAGTGTATCATTTTTGACGAATGTCACGTGCATGACG
CATCAGCAATGGCATTTAGATGTCTTCTTTCGGAGTATTCATACCCGGGAAAGATACTGAAGGTCTCAG
CGACACCCCCTGGACATGAAGTTGATTTCAAAACACAGAAGGAGGTGAAGGTCATTGTTGAAGAATCT
TTGTCATTCCAGCAGTTTGTCTCCAATCTCGGCACAGGTTGCAATAGCGATATTCTCAAGCATGGGGTCA
ATGTGTTGGTCTATGTCGCAAGTTACAATGAGGTTGACACACTAAGCAAATTGCTCACAGACAGGAGCT
TCAAAGTTTCAAAAGTCGATGGGCGAACTATGAAAATCGGCAATGTTGAGATACCAACGAGTGGCACT
CAGGCTAAACCGCATTTCGTGGTTGCAACAAATATCATTGAAAATGGAGTCACATTGGACATTGATGTG
GTTGTGGACTTCGGTTTGAAAGTCGTGCCTGTATTGGACATTGACAATCGTCTCGTTCGATACACGAAG
AAGAGCATCAGTTATGGAGAAAGGATTCAAAGATTGGGGCGAGTTGGTCGAAACAAACCAGGAGCAG
CACTTCGTATTGGATTTACAGAGAAAGGACTCACTCAAATACCTCCAATAATCGCAACAGAAGCAGCTT
TTCTATGTTTCACTTATGGTTTGCCAGTTATGACTAACGGTGTGTCAACGAGCCTACTAGCGATGTGCAC
TGTCAAGCAAGCACGGACGATGCAACAATTTGAATTATCCCCGTTCTACACAGTGGCGTTGGTTCGATT
TGACGGGACAATGCACCAGGAAATTTTTCGATTGCTCAAGAGCTATAGACTGCGTGACTCAGAGGTAA
TCTTAAACAAGTTGGCTATACCAAACAGCAACGTATGTGGGTGGATGAGTGTTCGTGACTATAAACGAC
AAGGCTGCAACTTGGACTTGGATGAGAACATTCGTGTACCGTTTTACGTGAAAGACATCCCTGAAACTT
TGCACGAGAGAATATGGCAAGTGGTAGAAACCCACAAATCTGATGCAGGATTTGGAAGGATCTGTAGT
TCGAGTGCGTGCAAAATTGCGTATACGTTACAGACAGACATCCACTCCATTCCTCGGACAATTAAAATC
ATTGACGCACTGTTGGAGCAAGAGAGAACAAAGCAAGCACACTTCAGAGCTATGACCAGTCAATCCTG
CTCAAGTTCAAATTTCTCTCTGTCAAGCATCACCTCAGCCATTCGCTCAAAATACGCCAAAGACCATACA
GAAGAAACATTGGTGTTCTCCAAATGGCGAAGTCTCAGTTGCTAGAATTCAAGAACCTGAACATTGAT
CCAAGTTATCCTGAACTTGTCCGCAACTTTGGCGCCTTAGAATGTGTGCACCATCAAACAAAGGAAGGA
GTTTCAAAGGCGCTACAACTTAAGGGGCATTGGAATAAGCGACTCATCACTCGTGACGCAACATTAATG
CTTGGAGTTCTTGGTGGGGGGGCATGGATGATTTTCAGTTATTTGAGGGATAGCTTCAAAGAAGAAGT
TGTTCACCAAGGCTTCAATCGTAGGCAAAGACAAAAATTGAAATTCAGGCAAGCACGAGATAACAGAA
TGGCCAGGGAAGTGTATGGTGACGATTCAACTATGGAGGACTACTTTGGTTCTGCATACTCAAAGAAA
GGAAAGAGCAAAGGAAAGACTAGAGGGATGGGAACGAAAACACGCAAATTTGTGAACATGTACGGG

TACGATCCCACAGACTATAACTTTGTTCGCTTTGTTGATCCATTGACTGGTCACACCCTGGACGAGAATC
CTCTTATGGACATCAACTTGGTGCAGGAACACTTCTCACAGATTCGCAATGATTACATCGGAGACGACA
AAATCACCATGCAGCACATAATGTCAAATCCAGGTATTGTCGCATACTATATCAAGGATGCAACGCAGA
AAGCCCTCAAAGTGGACCTTACTCCACACAACCCATTGCGTGTATGTGACAAAACTGCAACTATTGCAG
GATTTCCAGAGAGAGAGTTTGAATTGAGGCAGACAGGACACCCAATTTTTGTTGAACCTAATGCGATCC
CAAAGATCAATGAAGAGGGGGACGAAGAAGTTGACCACGAAAGTAAATCACTGTTCAGAGGCCTGAG
AGACTATAATCCAATCGCAAGCTCAATATGCCAATTGAATAACTCATCTGGTGCTAGACAAAGTGAAAT
GTTTGGACTTGGCTTTGGGGGTTTAATTGTCACGAATCAGCATTTGTTCAAAAGGAATGACGGAGAGCT
AACAATCCGATCGCATCATGGGGAATTCGTAGTGAAGGACACAAAAACTCTCAAACTGCTTCCTTGCAA
AGGTCGAGACATAGTGATCATCAGATTACCAAAGGATTTCCCTCCTTTTCCGAAGAGGTTGCAGTTCCG
CACCCCGACGACTGAGGACAGAGTTTGTTTAATTGGTTCAAATTTCCAAACGAAGAGCATTTCAAGCAC
CATGTCGGAAACAAGCGCAACATATCCAGTTGATAACAGTCATTTCTGGAAACACTGGATTAGCACGAA
GGATGGTCATTGCGGATTACCCATCGTGAGCACTCGAGATGGCAGTATTCTTGGGCTACACAGTCTTGC
AAATTCAACGAACACCCAGAATTTCTATGCAGCTTTCCCTGACAACTTCGAGACCACATACTTGTCAAAT
CAAGACAATGATAACTGGATAAAGCAGTGGCGGTACAACCCGGATGAAGTTTGCTGGG

(BamHI restriction site)

Fragment BamHI-NIb (1664bp)

GATCCCTACAACTCAAGAGGGACATTCCACAGAGTCCGTTTACAATTTGTAAACTGCTAACGGATCTTG
ATGGGGAATTTGTTTACACTCAGTCCAAAACTACACATTGGCTCAGAGATAGATTAGAAGGAAATTTGA
AAGCAGTTGGAGCCTGCCCTGGGCAGTTGGTTACTAAGCATGTCGTTAAAGGCAAATGTACACTCTTTG
AAACATACCTGTTGACTCATCCAGAGGAGCACGAATTCTTTCGACCTTTAATGGGAGCATACCAAAAGA
GTGCTCTAAATAAGGACGCATACGTCAAAGATCTGATGAAGTATTCAAAACCAATCGTCGTTGGTGCAG
TTGACTGTGATCAATTTGAACGTGCTGTTGATGTGGTCATTTCGATGCTAATTTCCAAAGGTTTTGAAGA
ATGTAATTACGTCACTGATCCAGATGACATATTCTCAGCACTTAACATGAAAGCAGCAGTTGGCGCTTT
GTACAGTGGAAAGAAAAGAGACTATTTTAAGAACGTGTCAGACCAGGACAAGGAAAGTTTCGTGCGA
GCTAGTTGCAAACGTTTGTTCATGGGAAAGAAAGGAGTGTGGAATGGCTCTTTGAAGGCAGAATTGCG
CCCTAAAGAGAAGGTAGAGGCTAATAAAACTCGATCATTCACAGCAGCACCGATTGATACCCTTCTGGG
GGGAAAAGTGTGTGTTGATGACTTCAATAATCAGTTTTACAGCCTGAATTTACATTGTCCATGGAGCGT
TGGGATGACAAAATTCAGAGGTGGTTGGGACAAACTGCTTAGAGCACTGCCAGAAGGATGGATTTACT
GTGATGCCGATGGCTCTCAATTTGACAGTTCCCTCTCACCGTACTTAATCAATGCAGTTCTCAATATTCGT
CTGGCATTTATGGAAGAATGGGACATTGGTGAACAAATGCTTTCAAACCTGTACACGGAGATTGTATAT
ACACCAATTGCTACACCAGATGGCACTATTGTTAAGAAGTTCAAGGGCAACAATAGTGGTCAACCCTCG
ACAGTTGTTGACAATACACTCATGGTTATTTTGGCAATGACATATTCACTCCTTAAGCTTGGATACCATC
CGGATACACACGATTGCATTTGTCGGTACTTCGTGAATGGTGATGATCTTGTCCTTGCAGTGCACCCAG
CATACGAGAGCATTTATGATGAGCTTCAAGAACACTTTTCCCAACTTGGATTGAATTACACATTCGCCAC
AAAGACTGAAAACAAGGAAGAGCTGTGGTTTATGTCACATAAAGGCGTTCTCTACGATGACATGTACA
TTCCTAAGCTAGAGCCTGAGAGGATTGTATCAATACTTGAATGGGACAGATCAAATGAGCCAATCCATC
GATTGGAGGCAATTTGTGCATCAATGGTGGAAGCGTGGGGTTATAAGGAGCTGCTGAGGGAGATCCG
GAAATTTTACAGTTGGGTTCTTGAACAAGCACCATACAATGCTCTTTCAAAAGATGGAAAAGCCCCGTA
CATTGCGGAGACAGCACTGAAGAAGCTTTACACTGACACTGAAGCATCTGAGACAGAAATTGAGCGAT
ATCTTGAAGCTTTTTACGACGACTTTAACGATGATGGTGAGTCCAACGTTGTTGTGCACCAAGCT
   N   V   V   V   H   Q   A

Fragment BamHI-NIb-GFP (2393 bp)

GATCCCTACAACTCAAGAGGGACATTCCACAGAGTCCGTTTACAATTTGTAAACTGCTAACGGATCTTG
ATGGGGAATTTGTTTACACTCAGTCCAAAACTACACATTGGCTCAGAGATAGATTAGAAGGAAATTTGA
AAGCAGTTGGAGCCTGCCCTGGGCAGTTGGTTACTAAGCATGTCGTTAAAGGCAAATGTACACTCTTTG
AAACATACCTGTTGACTCATCCAGAGGAGCACGAATTCTTTCGACCTTTAATGGGAGCATACCAAAAGA
GTGCTCTAAATAAGGACGCATACGTCAAAGATCTGATGAAGTATTCAAAACCAATCGTCGTTGGTGCAG
TTGACTGTGATCAATTTGAACGTGCTGTTGATGTGGTCATTTCGATGCTAATTTCCAAAGGTTTTGAAGA
ATGTAATTACGTCACTGATCCAGATGACATATTCTCAGCACTTAACATGAAAGCAGCAGTTGGCGCTTT
GTACAGTGGAAAGAAAAGAGACTATTTTAAGAACGTGTCAGACCAGGACAAGGAAAGTTTCGTGCGA
GCTAGTTGCAAACGTTTGTTCATGGGAAAGAAAGGAGTGTGGAATGGCTCTTTGAAGGCAGAATTGCG
CCCTAAAGAGAAGGTAGAGGCTAATAAAACTCGATCATTCACAGCAGCACCGATTGATACCCTTCTGGG
GGGAAAAGTGTGTGTTGATGACTTCAATAATCAGTTTTACAGCCTGAATTTACATTGTCCATGGAGCGT
TGGGATGACAAAATTCAGAGGTGGTTGGGACAAACTGCTTAGAGCACTGCCAGAAGGATGGATTTACT
GTGATGCCGATGGCTCTCAATTTGACAGTTCCCTCTCACCGTACTTAATCAATGCAGTTCTCAATATTCGT
CTGGCATTTATGGAAGAATGGGACATTGGTGAACAAATGCTTTCAAACCTGTACACGGAGATTGTATAT
ACACCAATTGCTACACCAGATGGCACTATTGTTAAGAAGTTCAAGGGCAACAATAGTGGTCAACCCTCG
ACAGTTGTTGACAATACACTCATGGTTATTTTGGCAATGACATATTCACTCCTTAAGCTTGGATACCATC
CGGATACACACGATTGCATTTGTCGGTACTTCGTGAATGGTGATGATCTTGTCCTTGCAGTGCACCCAG
CATACGAGAGCATTTATGATGAGCTTCAAGAACACTTTTCCCAACTTGGATTGAATTACACATTCGCCAC
AAAGACTGAAAACAAGGAAGAGCTGTGGTTTATGTCACATAAAGGCGTTCTCTACGATGACATGTACA
TTCCTAAGCTAGAGCCTGAGAGGATTGTATCAATACTTGAATGGGACAGATCAAATGAGCCAATCCATC
GATTGGAGGCAATTTGTGCATCAATGGTGGAAGCGTGGGGTTATAAGGAGCTGCTGAGGGAGATCCG
GAAATTTTACAGTTGGGTTCTTGAACAAGCACCATACAATGCTCTTTCAAAAGATGGAAAAGCCCCGTA
CATTGCGGAGACAGCACTGAAGAAGCTTTACACTGACACTGAAGCATCTGAGACAGAAATTGAGCGAT
ATCTTGAAGCTTTTTACGACGACTTTAACGATGATGGTGAGTCC**AACGTTGTTGTGCACCAAGCT**ATGG
TGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
TTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCAGCTACGGCGTG
CAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTT
CGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAA
CGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT
CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCC
GGGATCACTCACGGCATGGACGAGCTGTACAAGGGTACC**AACGTTGTTGTGCACCAAGCT**

        N   V   V   V   H   Q   A

Fragment PA11m without STOP and followed by a new cleavage site (NVVHQ/A) (2271 bp)
ATGAAACATTTGCTCCAATATTTCTTGCTCCTATTCTTAATCCCTAAAATCTGTTTTACCATCATCCCTGCC
GTTCACAGCCTCTGCACTAAAGACCAGCAACTATCATTGCTCCATTTGAAGAAAAGCCTTCAATTTTCTC
ATGATCCTGATTCTGATTCATACCCAACCAAGGTTATATCTTGGAATTCAAGCACCGATTGTTGTTCTTG
GCTTGGTGTTAATTGCAGTAGTGATGGGCATGTCGTTGGTCTTGACCTTAGCAGCGAAGCTATCAACGA
TGGCATTGACGATTCAAGCAGTCTCTTCGATCTTCAACACCTTCAAAGCCTCAATTTGGCTGACAACCAT
TTTACCTATGGTACTCGCATTCCATCTGCAATCGGAAAGCTTGTGAACTTGAGGTATCTAAATTTATCAT
CTTGCAGTTTCTATGGATCAATCCCAAAGTCAATAGCAAATCTAACACAATTGGTTAGTTTGCATTTGGG

ATTAAATACGTTCAGTGGTTCAATTGATTCTATTAGCTGGGAAAACCTTATTAATCTGGTAGACCTCCAG
ATGGATGACAACCTACTTGAGGGGAGTATTCCATCGTCTCTCTTTTATCTTCCCTTATTGACACAACTAGT
ACTTTCCCGCAATCAATTCTCTGGTAAACTTCATGCATTTTCTAACACCTCTTCCGACTTAGAATATTTGG
ACCTTTCAGAAAACCAGATTCAAGGCAAGATACCCCATTGGATTTGGAGTTTCAGTCATCTTTATTACCT
AAATCTTTCTTGCAACTCTTTGGTAACTCTAGAAGCTCCTTTATATAATTCTAGTGTATCAATAGTTGACC
TTCATTCAAACCAACTCCAGGGTCAAATCCCAACTTTCATACCATTTGGTTACCAGCTGGATTACTCAGG
CAACCATTTCAATTCTATACCATCTGACATTGGTTATTTCTTCACTTCCACAATGTTCTTCTCTCTTTCAAG
CAATAACTTGCATGGGCTCATTCCGGCATCAATATGCAATGCGACAAGTTTTCTTATGAGTCTTGATCTG
TCCAATAATTTTCTGAGTGGCATTATTCCCCCATGCTTGACTGCAATGCGCGGTCTCAGAGTACTTAATT
TAGCAAGAAACAACCTCACTGGAACTATTTCTAATTTTCAAGTTACTGAATATAGTTTATTAGAAATTCT
AAAGCTCGATGGAAATCAGTTAGGTGGTCAGTTTCCAAAATCTCTAGGTAACTGCATACAGTTACAGGT
TTTAAACTTGGGAAACAATCGTATAACAGATACATTTCCATGCTTGTTAAAAAACATGTCCACCTTGCGT
GTCCTTGTGTTGCGGTCCAACAACTTCTATGGAGGAATTGGATGTCCCAACACCTATGGCACCTGGCCA
GTGCTTCAAATCATACACCTAGCTCACAACAATTTCACTGGTGAAATACCGGGAATATTTTTGACAACAT
GGCAGGTAATGATGGCTCCCGAGGATGGTCCCCTATCGATTGTCAAATTCCAACTGGATACAATTATTG
CGGGAAAATCAATGTTGATTGATTATTCTTTTAATGATCGTATAACAGTTACCAGCAAAGGGTTAGAGA
TGGATCTAGTAAGGATTCTATCTATCTTCACCTTGATTGACTTCTCTTGCAACAACTTCAGTGGACCAATA
CCTAAGGAAATGGGAGAATTCAAATCACTACATGTCCTTAACTTGTCCAGAAATTCTTTGACAGGCGAA
ATCCCATCCTCATTTGGTAACATGCAGGTACTCGAGTCCTTGGACCTGTCACAGAACAAGTTGGGCGGG
GAAATTCCACAACAGTTGGCAAAGCTTACTTTCCTTTCGTTCTTGAATATCTCATATAATCAACTGGTCG
GCAGGATCCCACCCAGTACTCAGTTTTCAACATTTCCAAAAGACTCATTTACAGGAAACAAAGGACTAT
GGGGGCCTCCTTTGACAGTGGATAACAAAACAGGATTATCACCACCACCAGCATTAAATGGAAGCCTTC
CAAATTCTGGCCATCGTGGGATTAATTGGGATCTGATCAGTGTTGAAATTGGATTTACAGTTGGCTTTG
GAGCTTCCGTTGGGTCACTTGTGTTGTGCAAGAGATGGAGTAAGTGGTATTACAGAGCTATGTACAGG
ATGGTTCTTAAGATATTCCCACAGCTGGAGGAAAGAATTGGAATTCATCGAAGACATGTTCACATAAAT
CGAAGGTGGAGACGT**AACGTTGTTGTGCACCAAGCT**

<div style="color:red">   N   V   V   V   H   Q   A</div>

Note: the genetic code degeneracy has been used to modify the cleavage site.
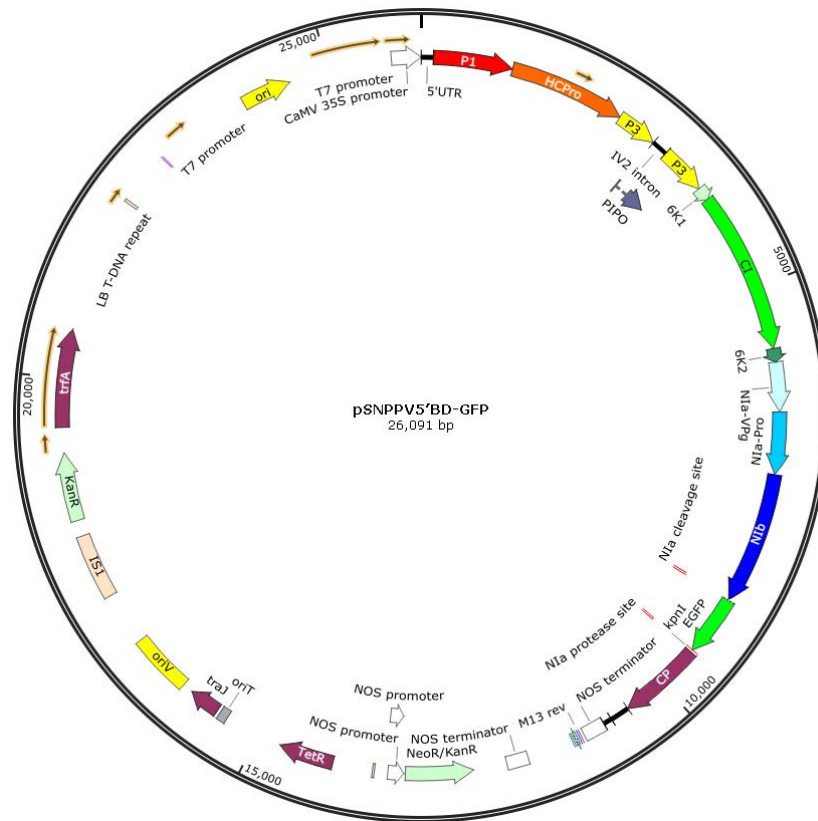
<u>Fragment CP-3'UTR-polyA-NOSt-XbaI (1576 bp)</u>

GCTGACGAAAGAGAAGACGAGGAGGAAGTTGATGCAGGCAAGCCGAGTGTAGTTACTGCACCGGCA
GCAACTAGCCCAATACTTCAACCACCTCCAGTCATACAGCCTGCACCCCGGACTACGGCGTCAATGCTC
AACCCCATTTTCACGCCAGCAACAACTCAACCAGCAACAAAACCAGTTTCACAGGTGTCAGGACCTCAA
CTGCAAACTTTTGGAACATATGGTAATGAGGATGCATCACCTAGCAACTCAAACGCGCTAGTCAACACA
AACAGAGACAGGGACGTCGATGCAGGATCAGTTGGAACTTTTACAGTGCCACGTTTGAAGGCAATGAC
TTCGAAACTATCTCTGCCAAAGGTGAAGGGAAAGGCTATTATGAACTTGAACCATTTGGCACATTATAG
TCCTGCACAGGTTGACTTGTCAAACACGAGAGCTCCGCAGTCTTGTTTCCAAACTTGGTATGAAGGAGT
TAAGCGAGATTATGATGTCACGGACGATGAAATGAGCATCATTTTAAATGGTCTTATGGTTTGGTGCAT
AGAGAATGGAACATCCCCGAATATCAATGGAATGTGGGTGATGATGGATGGGGAAACACAAGTGGAG
TATCCAATAAAGCCATTGTTGGATCATGCGAAACCCACTTTTAGACAAATTATGGCACATTTCAGTAACG
TGGCTGAAGCGTATATTGAAAAACGAAATTATGAAAAAGCATACATGCCAAGGTATGGAATTCAGCGC
AACCTGACAGACTACAGCCTCGCCAGATATGCCTTTGATTTTTACGAAATGACTTCAACGACACCAGTAC
GGGCACGTGAAGCTCATATCCAGATGAAGGCAGCAGCATTGAGAAATGTTCAAAATCGTTTATTTGGC
TTGGATGGAAACGTCGGAACACAAGAAGAGGACACAGAGAGACACACCGCTGGTGATGTTAATCGCA
ACATGCACAACCTCCTCGGTGTGAGGGGAGTGTAGTGGTCTCGGTATCTATCATAAACTCTACCTGGGT

GAGAGTCTAATCATCCAGTTGTTTTTAGATTCCTGTTAGCATCCTTTTCTCCGCTTTAATAGCAGTACATT
CAGTGAGGTTTTACCTCCATATGTTCTAGTCTGTTATTGTCGAACACAGGCCCTTGTATCTGATGTAGCG
AGTGCTTCACTCCATTCGGGTTATAGTTCTTGTGCAAGAGACAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAATGCATGCCTGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGC
CGGTCTTGCGATGATTATCATCTAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGC
ATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAA
ACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCT

Fragment GFP-CP-3'UTR-NOSt-XbaI

GCCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCG
ACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC
CCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCAGCTAC
GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT
GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCA
GAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC
GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGC
ACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC
CGCCGCCGGGATCACTCACGGCATGGACGAGCTGTACAAGGGTACC<mark>AACGTTGTTGTGCACCAA</mark>GCTG
ACGAAAGAGAAGACGAGGAGGAAGTTGATGCAGGCAAGCCGAGTGTAGTTACTGCACCGGCAGCAAC
TAGCCCAATACTTCAACCACCTCCAGTCATACAGCCTGCACCCCGGACTACGGCGTCAATGCTCAACCCC
ATTTTCACGCCAGCAACAACTCAACCAGCAACAAAACCAGTTTCACAGGTGTCAGGACCTCAACTGCAA
ACTTTTGGAACATATGGTAATGAGGATGCATCACCTAGCAACTCAAACGCGCTAGTCAACACAAACAGA
GACAGGGACGTCGATGCAGGATCAGTTGGAACTTTTACAGTGCCACGTTTGAAGGCAATGACTTCGAA
ACTATCTCTGCCAAAGGTGAAGGGAAAGGCTATTATGAACTTGAACCATTTGGCACATTATAGTCCTGC
ACAGGTTGACTTGTCAAACACGAGAGCTCCGCAGTCTTGTTTCCAAACTTGGTATGAAGGAGTTAAGCG
AGATTATGATGTCACGGACGATGAAATGAGCATCATTTTAAATGGTCTTATGGTTTGGTGCATAGAGAA
TGGAACATCCCCGAATATCAATGGAATGTGGGTGATGATGGATGGGGAAACACAAGTGGAGTATCCA
ATAAAGCCATTGTTGGATCATGCGAAACCCACTTTTAGACAAATTATGGCACATTTCAGTAACGTGGCT
GAAGCGTATATTGAAAAACGAAATTATGAAAAAGCATACATGCCAAGGTATGGAATTCAGCGCAACCT
GACAGACTACAGCCTCGCCAGATATGCCTTTGATTTTTACGAAATGACTTCAACGACACCAGTACGGGC
ACGTGAAGCTCATATCCAGATGAAGGCAGCAGCATTGAGAAATGTTCAAAATCGTTTATTTGGCTTGGA
TGGAAACGTCGGAACACAAGAAGAGGACACAGAGAGACACACCGCTGGTGATGTTAATCGCAACATG
CACAACCTCCTCGGTGTGAGGGGAGTGTAGTGGTCTCGGTATCTATCATAAACTCTACCTGGGTGAGA
GTCTAATCATCCAGTTGTTTTTAGATTCCTGTTAGCATCCTTTTCTCCGCTTTAATAGCAGTACATTCAGT
GAGGTTTTACCTCCATATGTTCTAGTCTGTTATTGTCGAACACAGGCCCTTGTATCTGATGTAGCGAGTG
CTTCACTCCATTCGGGTTATAGTTCTTGTGCAAGAGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAATGCATGCCTGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGT
CTTGCGATGATTATCATCTAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGA
CGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAACAA
AATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCT

Supplementary Figure 4.S1 Overview of plasmid pSNPPV5'BD-GFP based on *PPV*. Functional cistrons are marked in different colors.