Changes in Anthocyanin Production during Domestication of *Citrus*[1][OPEN]

Eugenio Butelli*, Andrés Garcia-Lor, Concetta Licciardello, Giuseppina Las Casas, Lionel Hill, Giuseppe Reforgiato Recupero, Manjunath L. Keremane, Chandrika Ramadugu, Robert Krueger, Qiang Xu, Xiuxin Deng, Anne-Laure Fanciullino, Yann Froelicher, Luis Navarro, and Cathie Martin

John Innes Centre, Norwich NR4 7UH, United Kingdom (E.B., L.H., C.M.); Instituto Valenciano de Investigaciones Agrarias, 46113 Moncada, Valencia, Spain (A.G.-L., L.N.); Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Centro di Ricerca per l'Agrumicoltura e le Colture Mediterranee, 95024 Acireale, Italy (C.L., G.R.-R.); Dipartimento di Agricoltura, Alimentazione e Ambiente, University of Catania, 95123 Catania, Italy (G.L.C.); United States Department of Agriculture-Agricultural Research Service National Clonal Germplasm Repository for Citrus and Dates, Riverside, California 92507-5437 (M.L.K., R.K.); University of California, Riverside, California 92521 (C.R.); Key Laboratory of Horticultural Plant Biology of the Ministry of Education, Huazhong Agricultural University, Wuhan 430070, People's Republic of China (Q.X., X.D.); Institut National de la Recherche Agronomique, UR1115 PSH, F-84914 Avignon, France (A.-L.F.); and CIRAD, Unité Mixte de Recherche AGAP, Station Institut National de la Recherche Agronomique, F-20230 San Giuliano, France (Y.F.)

ORCID IDs: 0000-0001-6397-277X (E.B.); 0000-0001-9147-9130 (L.H.); 0000-0001-7663-2299 (G.R.-R.); 0000-0002-8646-9293 (M.L.K.); 0000-0003-1786-9696 (Q.X.); 0000-0003-4490-4514 (X.D.); 0000-0001-5163-5960 (L.N.); 0000-0002-3640-5080 (C.M.).

Mandarin (*Citrus reticulata*), citron (*Citrus medica*), and pummelo (*Citrus maxima*) are important species of the genus *Citrus* and parents of the interspecific hybrids that constitute the most familiar commercial varieties of *Citrus*: sweet orange, sour orange, clementine, lemon, lime, and grapefruit. Citron produces anthocyanins in its young leaves and flowers, as do species in genera closely related to *Citrus*, but mandarins do not, and pummelo varieties that produce anthocyanins have not been reported. We investigated the activity of the *Ruby* gene, which encodes a MYB transcription factor controlling anthocyanin biosynthesis, in different accessions of a range of *Citrus* species and in domesticated cultivars. A white mutant of lemon lacks functional alleles of *Ruby*, demonstrating that *Ruby* plays an essential role in anthocyanin production in *Citrus*. Almost all the natural variation in pigmentation by anthocyanins in *Citrus* species can be explained by differences in activity of the *Ruby* gene, caused by point mutations and deletions and insertions of transposable elements. Comparison of the allelic constitution of *Ruby* in different species and cultivars also helps to clarify many of the taxonomic relationships in different species of *Citrus*, confirms the derivation of commercial varieties during domestication, elucidates the relationships within the subgenus *Papeda*, and allows a new genetic classification of mandarins.

Anthocyanins are phenolic compounds responsible for the red, blue, and purple colors in most angiosperm plants. They are accumulated in different tissues, particularly in flower petals, fruit peel, and leaves (Winkel-Shirley, 2001).

The pathway of anthocyanin biosynthesis is well characterized, and the structural genes that encode the core biosynthetic enzymes have been identified from many species (Tanaka et al., 2008). These genes are coordinately induced by a conserved regulatory complex formed by the interaction between MYB transcription factors, basic helix-loop-helix transcription factors, and WD40-repeat proteins, called the MBW complex (Quattrocchio et al., 1993; Koes et al., 2005; Ramsay and Glover, 2005). DNA-binding specificity is provided principally by the MYB factors. Mutations in the MYB regulatory genes often result in a lack of expression of the biosynthetic genes and loss of anthocyanin pigmentation, as exemplified by the occurrence of red and white varieties of grape (*Vitis vinifera*; Kobayashi et al., 2004), apple (*Malus domestica*; Espley et al., 2007), and species of *Petunia* (Quattrocchio et al., 1999; Hoballah et al., 2007) and *Antirrhinum* (Schwinn et al., 2006).

In *Citrus*, the ability to accumulate anthocyanins is not a universal feature (Fig. 1). In lemons and limes, purple pigmentation is clearly present in young leaves and flowers (new growth). In contrast, other cultivated *Citrus* varieties of economic importance are completely unable to synthesize anthocyanins (Hodgson, 1967). Sweet oranges, with the notable exception of the striking fruit-specific, red-colored, blood varieties, do not display any anthocyanin pigmentation.

The taxonomy of *Citrus*, a genus of plants belonging to the family Rutaceae (Supplemental Fig. S1), is complicated, as illustrated by the two most widely accepted classification systems, those proposed by Swingle and Tanaka, recognizing 16 and 162 species, respectively (Swingle, 1946; Swingle and Reece, 1967; Tanaka, 1977). In *Citrus*, the definition of a species is obscured by the ability of *Citrus* species to hybridize easily within the genus and also with closely related genera. Sexual promiscuity and thousands of years of human cultivation

have generated many hybrids whose origin cannot be inferred solely on the basis of morphological and geographical data. However, a growing body of evidence supports the view that there are three basic or primary species: mandarin (*Citrus reticulata*), pummelo (*Citrus maxima*), and citron (*Citrus medica*; Scora, 1975; Barrett and Rhodes, 1976). All the other so-called species are hybrids derived from these three primary species (Moore, 2001). A fourth species, belonging to the subgenus *Papeda*, was involved in the derivation of some limes (Nicolosi et al., 2000). The term secondary species has been used widely to define *Citrus* hybrids, since most of them are apomictic, polyembryonic, and normally propagated vegetatively by grafting. This ensures that plants maintain their hybrid genetic constitution over generations but also makes genetic studies challenging and crop improvement by conventional breeding almost impossible. Within a given species or hybrid, phenotypic differences between cultivars usually are the result of spontaneous or induced somatic mutations that have been selected and propagated.

We have shown previously that, in different blood varieties of sweet orange, the insertion of retrotransposons in the promoter of *Ruby*, a regulatory gene encoding a MYB transcription factor, is responsible for the fruit-specific accumulation of anthocyanins (Butelli et al., 2012). Here, our objective was to determine whether *Ruby* controls the accumulation of anthocyanins in young leaves and flowers of pigmented *Citrus* species and hybrids and whether mutations in this gene can account for the inability of some *Citrus* varieties to develop pigmentation. We used sequence information to determine whether the phylogeny of *Ruby* is congruent with the generally accepted phylogeny of *Citrus* and the proposed ancestry of commercially important hybrids.

[1] This work was supported by the Institute Strategic Program Understanding and Exploiting Plant and Microbial Secondary Metabolism from the Biotechnology and Biological Sciences Research Council (grant no. BB/J004596/1 to C.M. and E.B.), the European Union FP7 ATHENA collaborative project (grant no. 245121 to C.M., E.B., C.L., G.L.C, and G.R.R.), the Ministry of Economia y Competitividad (grant no. AGL2011-26490 to A.-G.L. and L.N.), the Ministry of Economía, Industria y Competitividad-Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria Fondo Europeo de Desarrollo Regional (grant no. RTA2015-00069-00-00 to A.-G.L. and L.N.), and the networking activities within the European Union-funded COST ACTION FA1106 QualityFruit.

* Address correspondence to eugenio.butelli@jic.ac.uk.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Eugenio Butelli (eugenio.butelli@jic.ac.uk).

E.B. and C.M. planned and designed the research; E.B., C.M., A.G.-L., C.L., G.L.C., and L.H. performed experiments; M.L.K., C.R., R.K., G.R.-R., A.-L.F., Y.F., L.N., Q.X., and X.D. provided plant material, DNA, information, and images; E.B. and C.M. wrote the article with input and comments from all the other authors.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.16.01701

RESULTS

The *Ruby* Locus in *Citrus* and Related Genera

Most taxonomists divide the group recognized as true *Citrus* fruit trees into six related genera: *Citrus*, *Poncirus*, *Microcitrus*, *Eremocitrus*, *Fortunella*, and *Clymenia* (Swingle and Reece, 1967; Supplemental Fig. S1). The genus *Citrus* is further divided into two subgenera: *Citrus*, which includes the common cultivated plants with commercially important fruit, and *Papeda* (Swingle, 1946; Krueger and Navarro, 2007).

We attempted a comprehensive analysis of the sequence variation at the *Ruby* locus in the three primary *Citrus* species, in four members of the subgenus *Papeda*, in several interspecific hybrids, and in accessions belonging to four genera related to *Citrus* that are relevant to understanding the regulation of anthocyanin production. For each, we sequenced the *Ruby* gene and at least 1.6 kb of the region upstream of the coding sequence. The species and hybrids we have studied are listed in Table I.

Ruby is involved in the fruit-specific accumulation of anthocyanins in blood varieties of sweet orange (Butelli et al., 2012). A second gene, which we named *Ruby2*, also encodes a putative R2R3 MYB transcriptional activator with the potential to regulate anthocyanin biosynthesis. The *Ruby2* and *Ruby* genes are arranged in tandem on chromosome 6 and are separated by an intergenic region ranging from 1.6 to 12 kb in different species. In *Citrus,* all the *Ruby2* alleles are predicted to encode nonfunctional proteins because of the presence of different combinations of deletions, frame shifts, and stop mutations (Supplemental Data Set S1).

No other genes with similarity to *Ruby* and encoding proteins belonging to subgroup 6 of the R2R3 MYB family (known to regulate anthocyanin biosynthesis; Stracke et al., 2001) were detected in the high-quality reference haploid clementine genome or in two different genome annotation projects of sweet orange (https://phytozome.jgi.doe.gov and http://citrus.hzau.edu.cn/orange/). A previous genome-wide targeted analysis of *R2R3 MYB* genes in sweet orange (Liu et al., 2014) and examination of the genome sequences of other *Citrus* species (Q. Xu and X. Deng, unpublished data) also failed to identify other paralogs of known anthocyanin-related *MYB* genes.

Based on these observations, the hypothesis was developed that, in all the accessions belonging to the genus *Citrus*, Ruby is the only R2R3 MYB transcriptional activator involved in anthocyanin biosynthesis. Our objective was to determine whether variation within the *Ruby* locus is responsible for the differential ability of different accessions to produce anthocyanins.

Ruby in Citron

Citron (*C. medica*) is a true species with a very low level of heterozygosity. Most varieties are able to accumulate

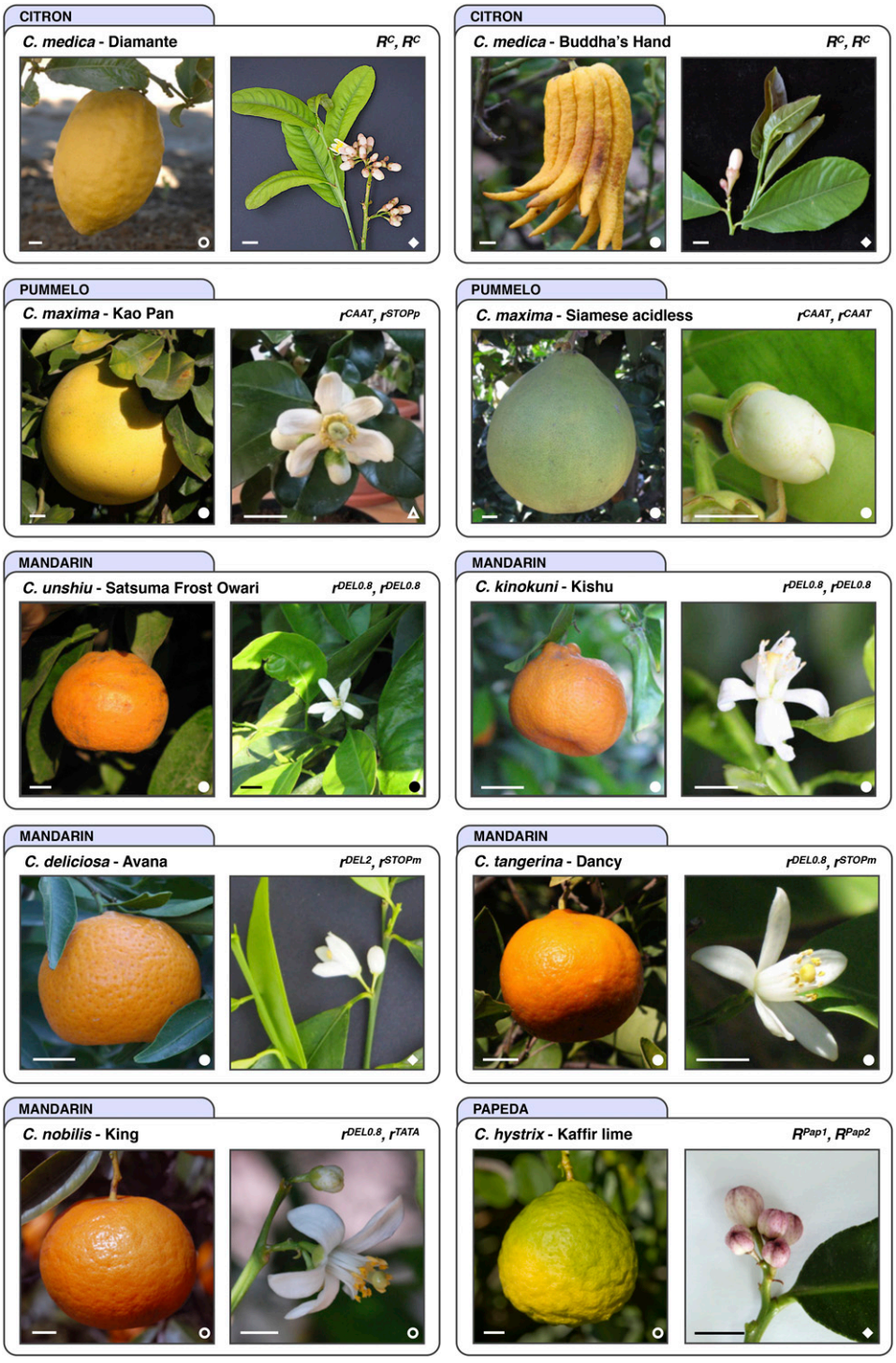

Figure 1. (*Figure continues on following page.*)

anthocyanins in young leaves and flowers (Fig. 1A), where cyanidin 3-glucoside and peonidin 3-(6′′-malonyl)-glucoside accounted for approximately 80% of the total anthocyanin content (Fabroni et al., 2016).

The three varieties of citron we tested (Table I) contained the same *Ruby* allele, predicted to be active and named R^C, in homozygous form. R^C was inherited, identical in sequence, in many hybrids, including lemon, where its expression was characterized.

Lying at position −902 relative to the start of transcription in R^C is a 5,435-bp retrotransposon that we named Tcl3 (Supplemental Fig. S2, A and B). Tcl3 is 92% and 91% identical in sequence to the two retroelements that are responsible for anthocyanin accumulation in Mediterranean (Tcs1) and Chinese (Tcs2) blood oranges, respectively (Butelli et al., 2012). All three elements are members of the same family of Copia-like long terminal repeat (LTR) retrotransposons, but while

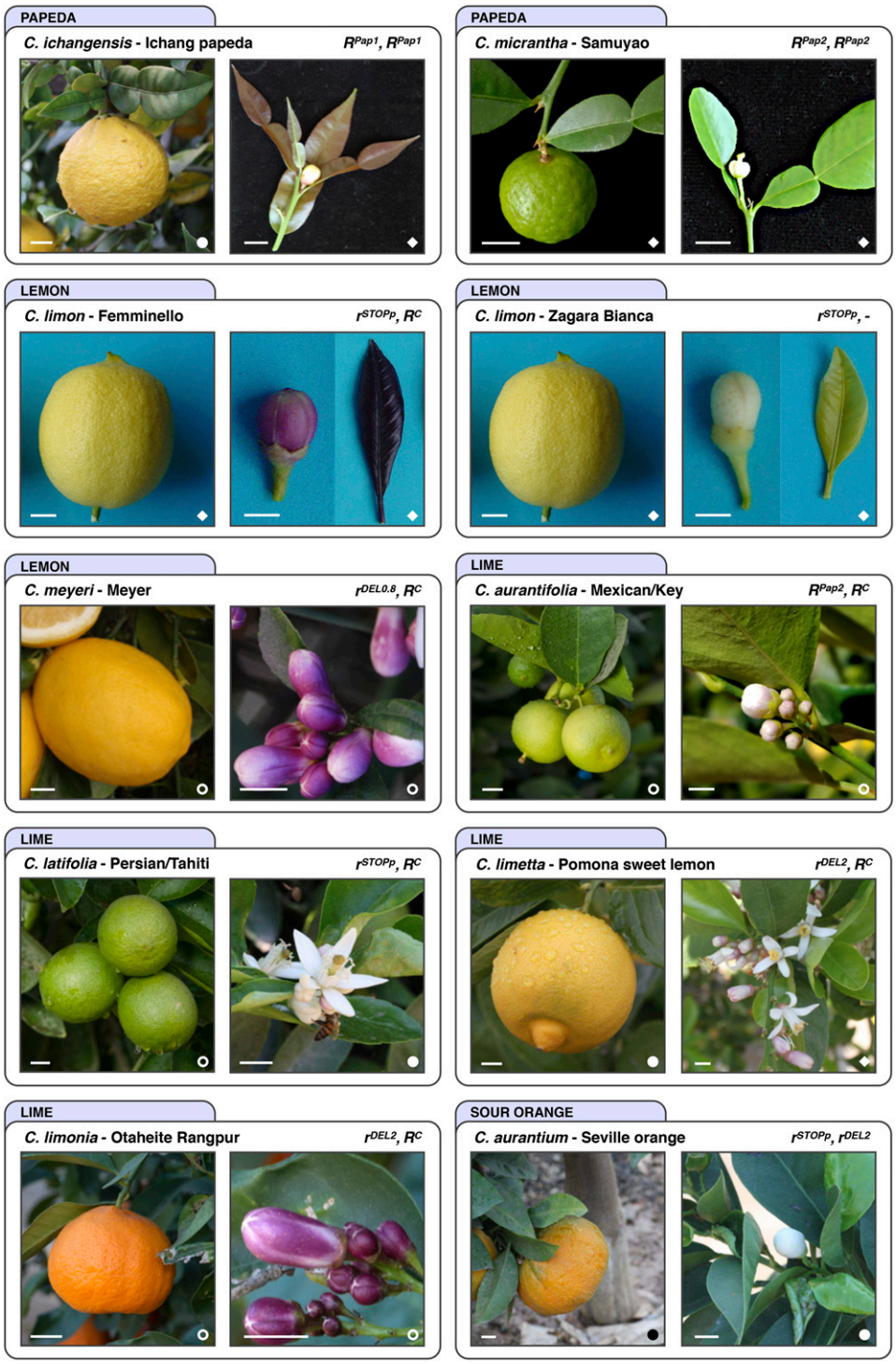

Figure 1. (*Figure continues on following page.*)

Tcs1 and Tcs2 are intact, active, and recently inserted elements, Tcl3 in citron is defective and is the result of an ancient retrotransposition event, as deduced by the divergence of LTR sequences that can be used to estimate the time of integration of LTR retrotransposons (SanMiguel et al., 1998). The two LTRs of Tcl3 differ by seven nucleotides and a 10-bp InDel, and the open reading frame of Tcl3 contains four stop mutations and one 5-bp deletion causing a frame-shift mutation that would inactivate the encoded polyprotein (Supplemental Fig. S2, B and C). These observations, together with its insertion at a considerable distance from the start of transcription, suggest that the presence of the Tcl3 retrotransposon likely has little or no impact on *Ruby* expression.

Figure 1. Phenotypes of some of the accessions considered in this study and listed in Table I. For each accession, the common and the scientific name according to Tanka's classification are provided, together with the allelic constitution at the *Ruby* locus. Images of fruit and flowers or young leaves are presented to show the presence or absence of anthocyanin pigmentation. The sources of the images are indicated by the following symbols: diamonds, pictures taken by the authors; open circles, UC-Riverside Citrus Variety Collection (http://www.citrusvariety.ucr.edu/index.html); closed circles, Citrus ID Tools (http://idtools.org/id/citrus/citrusid/); open triangles, Radoslav Krchnavek; closed triangles; Eve and George DeLange. Bars = ~1 cm.

Ruby in Pummelo

Pummelo (*C. maxima*) is a true species of *Citrus* with many cultivars and high genetic diversity. Anthocyanin pigmentation has never been reported in pummelo. The four accessions we initially considered (Chandler, Reinking, Siamese acidless, and Kao Pan) contain different combinations of three *Ruby* alleles that we named r^{STOPp}, r^{TATA}, and r^{CAAT} (Table I).

The r^{STOPp} allele is characterized by the presence of a stop mutation (TGG to TGA at positions equivalent to +1,481 in the R^C reference allele) in the third exon that would result in the truncation of the Ruby protein involving loss of its C-terminal 116 amino acids. This mutation results in a complete loss of function of the Ruby protein, as demonstrated by the lack of anthocyanin production when a truncated cDNA was transiently

Table I. *Species and hybrids used in this study*

Common name	Swingle	Tanaka	*Ruby* alleles			Source	Accession
CITRON							
Diamante	*C. medica*	*C. medica*	R^C	R^C		B	Palazzelli Certif.
Corsican	*C. medica*	*C. medica*	R^C	R^C		D	SRA613
Poncire commun	*C. medica*	*C. medica*	R^C	R^C		D	SRA701
Buddha's Hand	*C. medica*	*C. medica*	R^C	R^C		D	SRA 640
PUMMELO							
Chandler	*C. maxima*	*C. maxima*	r^{TATA}	r^{CAAT}		B	1A-F7P13
Reinking	*C. maxima*	*C. maxima*	r^{TATA}	r^{STOPp}		B	1A-F9P16
Purple-skin	*C. maxima*	*C. maxima*	r^{TATA}	R^P		E	-
Kao Pan	*C. maxima*	*C. maxima*	r^{CAAT}	r^{STOPp}		E	-
Siamese acidless	*C. maxima*	*C. maxima*	r^{CAAT}	r^{CAAT}		E	-
MANDARIN							
Satsuma Frost Owari	*C. reticulata*	*C. unshiu*	$r^{DEL0.8}$	$r^{DEL0.8}$		A	IVIA175
Satsuma Okitsu	*C. reticulata*	*C. unshiu*	$r^{DEL0.8}$	$r^{DEL0.8}$		A	IVIA195
Kishu	*C. reticulata*	*C. kinokuni*	$r^{DEL0.8}$	$r^{DEL0.8}$		A	IVIA678
Sunki	*C. reticulata*	*C. sunki*	r^{DEL2}	r^{DEL2}		A	IVIA239
Cleopatra	*C. reticulata*	*C. reshni*	r^{DEL2}	r^{DEL2}		A	IVIA385
Ponkan	*C. reticulata*	*C. reticulata*	r^{DEL2}	r^{STOPm}		A,B	IVIA482; A-F10P2
Avana	*C. reticulata*	*C. deliciosa*	r^{DEL2}	r^{STOPm}		A,B	IVIA189; A-F21P3
Dancy	*C. reticulata*	*C. tangerina*	$r^{DEL0.8}$	r^{STOPm}		A	IVIA434
Fuzhu	*C. reticulata*	*C. erythrosa*	$r^{DEL0.8}$	r^{STOPm}		A	IVIA571
Shekwasha	*C. reticulata*	*C. depressa*	$r^{DEL0.8}$	r^{DEL2}		A	IVIA238
King	*C. reticulata*	*C. nobilis*	$r^{DEL0.8}$	r^{TATA}		A	IVIA477
Kunembo	*C. reticulata*	*C. nobilis*	r^{DEL2}	$r^{DEL0.8}$		C	CRC3346
Keraji	*C. reticulata*	*C. keraji*	$r^{DEL0.8}$	$r^{DEL0.8}$		C	CRC3144
Oroval clementine	*C. reticulata*	*C. clementina*	r^{DEL2}	r^{STOPm}		B	Acireale Certif.
Nasnaran	*C. reticulata*	*C. amblycarpa*	R^{Pap2}	R^{Pap2}		A	IVIA478
Tachibana orange	*C. tachibana*	*C. tachibana*	$r^{DEL0.8}$	$r^{DEL0.8}$		A,C	IVIA237; CRC3150
Indian wild orange	*C. indica*	*C. indica*	R^I	R^I		A,C	IVIA550; CRC3163
PAPEDA							
Kaffir lime	*C. hystrix*	*C. hystrix*	R^{Pap1}	R^{Pap2}		A	IVIA178
Melanesian papeda	*C. macroptera*	*C. macroptera*	R^{Pap1}	R^{Pap2}		A	IVIA279
Ichang papeda	*C. ichangensis*	*C. ichangensis*	R^{Pap1}	R^{Pap1}		A	IVIA235
Samuyao	*C. micrantha* var *microcarpa*	*C. micrantha*	R^{Pap2}	R^{Pap2}		A	IVIA626
LEMON							
Femminello	*C. limon*	*C. limon*	r^{STOPp}	R^C		B	C-F8P2
Politi	*C. limon*	*C. limon*	r^{STOPp}	R^C		B	C-F3P1
Zagara Bianca	*C. limon*	*C. limon*	r^{STOPp}	-		B	B-F7P3
Dulce*	*C. limon*	*C. limon*	r^{STOPp}	R^C	r^{DEL2}	H	IVIA443
Meyer	*C. limon* hybrid	*C. meyeri*	$r^{DEL0.8}$	R^C		H	-
LIME							
Mexican/Key	*C. aurantifolia*	*C. aurantifolia*	R^{Pap2}	R^C		A	IVIA164
Persian/Tahiti	*C. aurantifolia*	*C. latifolia*	r^{STOPp}	R^C		G	-
Palestine sweet lime	*C. aurantifolia*	*C. limettioides*	$r^{DEL0.8}$	R^C		C	CRC1482
Mary Ellen	*C. aurantifolia*	*C. limettioides*	$r^{DEL0.8}$	R^C		C	CRC4053
Soh Synteng	*C. limon*	*C. limon*	$r^{DEL0.8}$	R^C		C	CRC3261
Limonette de Marrakech	*C. limon*	*C. limetta*	r^{DEL2}	R^C		C,D	CRC3989; SRA829
Limetta Dolce	*C. limon*	*C. limetta*	r^{DEL2}	R^C		B	C-F2P7
Pomona sweet lemon	*C. limon*	*C. limetta*	r^{DEL2}	R^C		C	CRC4068
Rangpur Kirumakki	*C. limon*	*C. limonia*	r^{DEL2}	R^C		C	CRC4131
Rangpur Otaheite	*C. limon*	*C. limonia*	r^{DEL2}	R^C		C	CRC2709
SOUR ORANGE							
Seville orange	*C. auratium*	*C. auratium*	r^{STOPp}	r^{DEL2}		G	-
GRAPEFRUIT							
Marsh	*C. paradisi*	*C. paradisi*	r^{STOPp}	r^{DEL2}		G	-
SWEET ORANGE							
Navel	*C. sinensis*	*C. sinensis*	r^{TATA}	r^{DEL2}		B	2B-F1-P14
Valencia	*C. sinensis*	*C. sinensis*	r^{TATA}	r^{DEL2}		B	6A2-F5P1
Tarocco blood	*C. sinensis*	*C. sinensis*	$R^{D.1}$	r^{DEL2}		B	4-F20P9
Moro blood	*C. sinensis*	*C. sinensis*	$R^{D.2}$	r^{DEL2}		B	8-F1P2
Jingxian blood	*C. sinensis*	*C. sinensis*	$R^{D.3}$	r^{DEL2}		F	-

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

Citrus-Related Genera						
PONCIRUS						
Trifoliate orange	*P. trifoliata*	*P. trifoliata*	R^{Pon}	R^{Pon}	B	-
MICROCITRUS						
Australian finger lime	*M. australasica* 'green fruit'	*M. australasica*	R^{Mic1}	R^{Mic1}	C	CRC3661
Australian finger lime	*M. australasica* var Sanguinea	*M. australasica*	R^{Mic2}	R^{Mic1}	B	C-F5P3
FORTUNELLA						
Nagami oval kumquat	*F. margarita*	*F. margarita*	r^{For}	r^{For}	H	-
Malayan kumquat	*F. polyandra*	*F. polyandra*	r^{For}	r^{For}	A	IVIA375
Calamondin	*Citrofortunella microcarpa*	*C. madurensis*	r^{For}	$r^{DEL0.8}$	H	-
SEVERINIA						
Chinese box orange	*S. buxifolia*	*S. buxifolia*	R^{Sev}	R^{Sev}	A	IVIA147

Citrus accessions were obtained from the following sources: A, Instituto Valenciano de Investigaciones Agrarias, Moncada, Valencia, Spain; B, Centro di Ricerca per l'Agrumicoltura e le Colture Mediterranee, Acireale, Italy; C, United States Department of Agriculture-Agricultural Research Service National Clonal Germplasm Repository for Citrus and Dates, Riverside, California; D, CRB CITRUS, Institut National de la Recherche Agronomique-CIRAD, Citrus Biological Resource Center, San Giuliano, Corsica, France; E, Huazhong Agricultural University, Wuhan, China; F, Jingzhou, Hunan Province, China; G, Waitrose Supermarket, Norwich, United Kingdom; H, Notcutts Garden Centre, Norwich, United Kingdom. In accessions of the genus *Citrus*, the *Ruby* alleles are color coded as follows: yellow, alleles derived from the mandarin pool; pink, alleles in mandarin introgressed from pummelo; red, alleles derived from pummelo; green, alleles derived from citron; olive green, alleles derived from *Papeda*; the unique allele in *C. indica* is in blue; the alleles in genera related to *Citrus* are not colored. The lemon accession indicated with an asterisk (Dulce) is a graft chimera and contains a third *Ruby* allele that may have originated from an acidless limetta. Photographs of fruit and flowers of the accessions indicated in boldface are provided in Figure 1.

overexpressed in *Nicotiana benthamiana* (Fig. 2B). In contrast, the full-length *Ruby* cDNA induced the accumulation of visible levels of delphinidin-3-rutinoside in *N. benthamiana* (Fig. 2A).

Although the coding sequences of the other two alleles identified in pummelo encoded intact proteins, we detected a number of changes in the core promoter that might disrupt *Ruby* expression. The r^{TATA} allele contains a T-to-C change in the center of the putative TATA box. This mutation lies at −32 bp relative to the start of transcription (as determined in lemon; see below). In the allele r^{CAAT}, we identified another T-to-C transition, this time within a putative CAAT box (at position −73 bp relative to the start of transcription and also present in r^{STOPp}). TATA and CAAT boxes are cis-acting elements that determine the efficiency of the promoter in many eukaryotic genes (Cooper and Hausman, 2016), and the two mutations, specific for the pummelo accessions, lie in a region that is well conserved in *Citrus* and related genera (Supplemental Fig. S3A). Sweet orange carries the r^{TATA} allele from its pummelo parent (see below; Butelli et al., 2012), and retrotransposon insertions into this allele provide either a new TATA box or activate a cryptic TATA box to activate *Ruby* expression in blood oranges (Butelli et al., 2012).

To confirm the prediction that both r^{CAAT} and r^{TATA} alleles are not expressed, we analyzed Chandler pummelo, since both alleles are present in this variety (Table I; Supplemental Fig. S3B). We were unable to isolate *Ruby* cDNA fragments, and we could detect no *Ruby* expression in leaves of Chandler pummelo (Supplemental Fig. S12C).

Although, theoretically, the lack of expression of *Ruby* from the r^{TATA} and r^{CAAT} alleles could have resulted from mutations in a closely linked regulator of *Ruby*, these alleles failed to complement alleles of *Ruby* with interruptions in the *Ruby* coding sequence (such as r^{STOPp}) in pummelo, and r^{TATA} failed to complement the alleles from mandarin (which carry large deletions in the *Ruby* gene) when combined in sweet orange (Butelli et al., 2012).

All four pummelo varieties analyzed contained at least one copy of a *Ruby* gene with an intact coding sequence. Therefore, pummelo appears to maintain the potential to produce anthocyanins, a potential that could be fulfilled by recombination of sequences lying between the mutated CAAT box allele and the mutated TATA box allele. Such an event is possible because pummelo, unlike the majority of cultivated *Citrus* plants, produces true zygotic, monoembryonic seeds. Although purple pummelo varieties have not been reported, an individual wild tree bearing purple/red-skinned fruit was identified by Xiuxin Deng in a mountainous region of Hubei province, China (Supplemental Fig. S3C). Sequence analysis indicated that this pigmented variety carried a *Ruby* allele with both an active CAAT box and an active TATA box that might be the result of such a recombination event. This unique allele, named R^{P}, was predicted to be fully active, since an intact coding sequence was under the

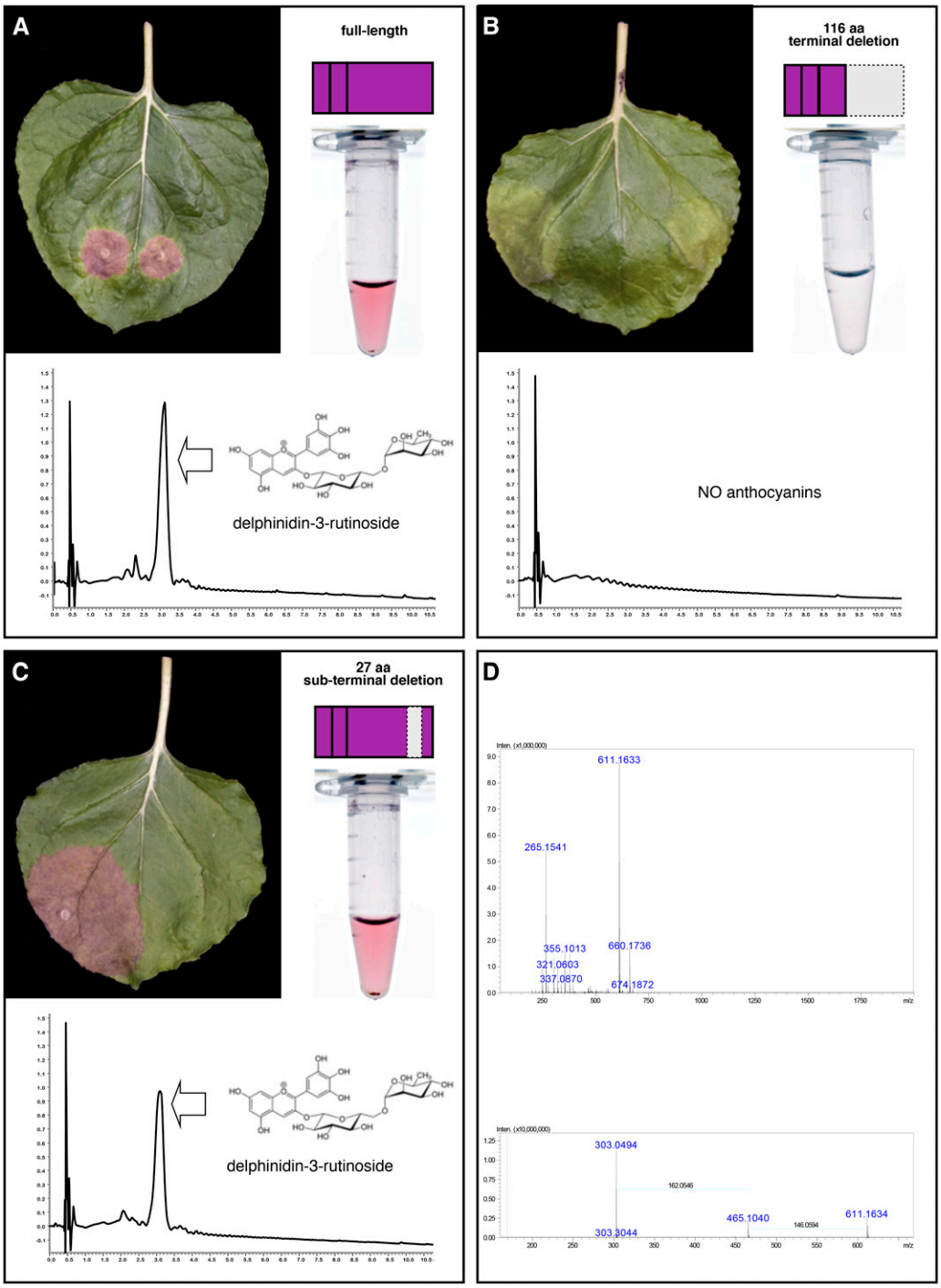

Figure 2. Functionality of different *Ruby* alleles. *N. benthamiana* leaves were infiltrated with cDNA constructs corresponding to three different *Ruby* alleles. The phenotypes of leaves, the color of the leaf extracts, and the HPLC scans at 525 nm 7 d after infiltration are shown. A, Transient overexpression of *Ruby* cDNA encoding a full-length protein isolated from the flesh of Moro blood orange. B, Transient overexpression of a cDNA construct encoding a protein with a terminal deletion of 116 amino acids corresponding to the alleles found in some pummelo and mandarin accessions and hybrids derived from them. C, Transient overexpression of *Ruby* cDNA isolated from the fruit peel of *M. australasica* var Sanguinea encoding a protein with a 27-amino acid (aa) subterminal deletion. D, Evidence of the identity of delphinidin-3-rutinoside in sample C by positive mode electrospray MS1 and MS2. For this compound, the expected mass is 611.1607 (found mass, 611.1633; error, 4.25 ppm). Note the fragment representing delphinidin (expected mass, 303.0499) and the losses of a hexose moiety (expected loss, 162.0528) and a rhamnose moiety (expected loss, 146.0579).

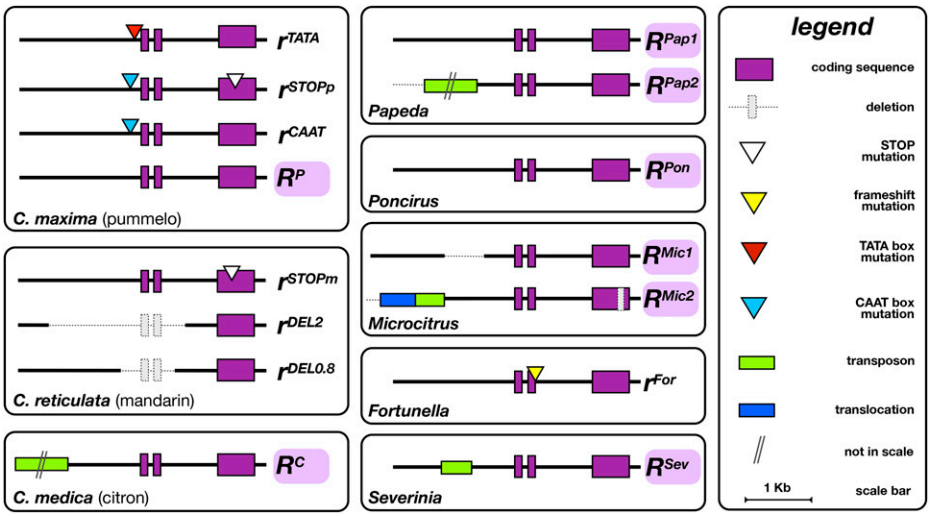

Figure 3. Schematic representation of different *Ruby* alleles identified in primary species of *Citrus* and related genera. Functional alleles (*R*) associated with anthocyanin pigmentation are highlighted in purple.

control of a core promoter without mutations in the CAAT or TATA box (Fig. 3; Supplemental Fig. S3). Although there were mismatches in the sequence of the region upstream of the TATA box in the R^P alleles, this sequence bore a greater similarity to the equivalent regions of r^{TATA} than to r^{CAAT}, including the absence of a 22-bp deletion. Downstream of the active TATA box motif in R^P, the sequence differed from r^{TATA}, and most of the single-nucleotide polymorphisms present in the area sequenced were identical between R^P and r^{CAAT} (Supplemental Fig. S4). This new active R^P allele supports the possibility that it might be the result of recombination. Although the occurrence of such a recombination event is likely to be rare, this particular individual might have been identified and preserved by the local population because of its colored fruit peel. The phenotype of the R^P allele might reflect the expression pattern of *Ruby* in a pigmented ancestor of pummelo, before loss-of-function *Ruby* alleles began to predominate as a result of evolution or human selection.

Ruby in Mandarin

The mandarin group is composed of numerous varieties with considerable genetic diversity but is treated as a single species under the name *C. reticulata* according to Swingle (Garcia-Lor et al., 2015).

We analyzed two common types of mandarin: Ponkan, an old variety thought to be a pure, prototypical *C. reticulata* (Hodgson, 1967), and Avana, a Mediterranean variety also known as Willowleaf. The structure and the sequence of the *Ruby* locus was identical in the two varieties and carried one allele (r^{STOPm}) with the same stop mutation identified in pummelo and another (r^{DEL2}) with a 2-kb deletion involving the first two exons and a large portion of the upstream regulatory region of *Ruby* (Table I; Fig. 3; Supplemental Fig. S5). Sequence alignments suggested that the r^{STOPm} allele was derived from pummelo because the two alleles, r^{STOPm} and r^{STOPp}, showed 99% nucleotide identity and contained identical polymorphisms, including the distinctive stop mutation, that were not present in *Ruby* alleles from other *Citrus* or related species (Supplemental Fig. S5). Comparative sequencing of different *Citrus* varieties (Wu et al., 2014) has indicated a substantial admixture from pummelo in Ponkan and Willowleaf mandarins. In particular, the reported nucleotide heterozygosity in the terminal region of chromosome 6, where *Ruby* is located, supports the likely pummelo origin of a *Ruby* allele in some varieties of mandarin.

Since an ancestral mandarin may have had a distinct allelic composition at the *Ruby* locus, we looked among existing mandarins for different alleles of *Ruby*.

Tanaka (1977) recognized 36 mandarin species separated into five taxonomic groups. We sequenced the *Ruby* alleles from 17 mandarin varieties, representative of accessions from all five groups (Table I).

On the basis of their constitution of *Ruby* alleles, these mandarin accessions could be divided into three groups (Fig. 4). Group A ($r^{DEL0.8}$ and $r^{DEL0.8}$) was characterized by being homozygous for a deletion of ~0.8 kb that spans the first two exons of the *Ruby* gene (between positions −175 and +629 relative to the start of transcription in the reference allele R^C). Mandarins of group B (r^{DEL2} and r^{DEL2}) were homozygous for a larger deletion of ~2 kb (upstream of position +729 in R^C). Mandarins of group C (r^{DEL2} and r^{STOPm}) were polymorphic for *Ruby* alleles, containing one allele introgressed from pummelo (r^{STOPm}) and one allele with the 2-kb deletion (r^{DEL2}) also present in group B. This polymorphic genetic constitution has probably been maintained because group C varieties, including Ponkan and Willowleaf, are highly apomictic. None of the alleles in mandarin encode an active *Ruby* protein, explaining why mandarins have never been associated with anthocyanin pigmentation.

The short deletion allele ($r^{DEL0.8}$) is highly polymorphic (Supplemental Fig. S5). Group A includes varieties

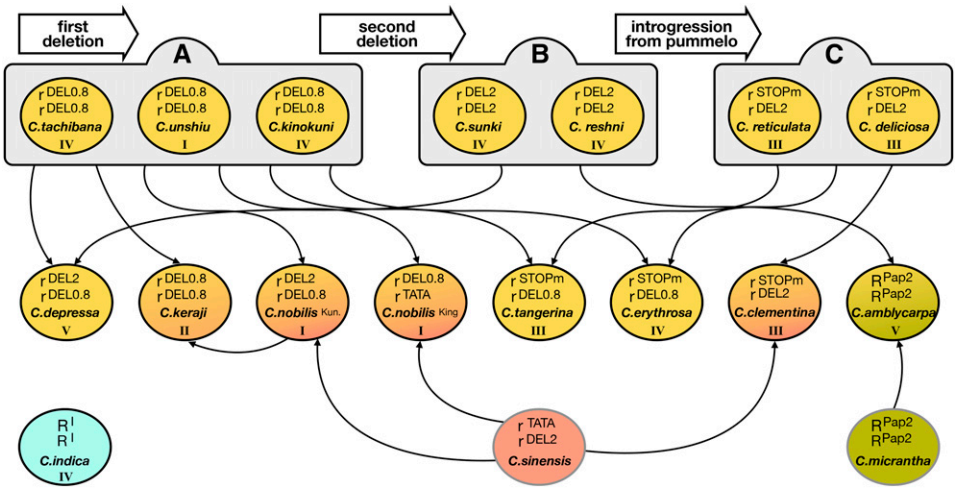

Figure 4. Classification of mandarins based on the analysis of the *Ruby* gene. Mandarins can be classified into three groups (A, B, and C) on the basis of two deletions and an introgression from pummelo. Many mandarin varieties are hybrids between members of these three groups or interspecific hybrids involving other *Citrus* species. *C. indica*, recognized by both Tanaka and Swingle as a mandarin species, is not related to mandarins. Roman numbers indicate the five taxonomic groups recognized by Tanaka (Hodgson, 1967). The gradation in color shading indicates the contribution from genomes of other species. Common names for the mandarin accessions are provided in Table I.

of Japanese origin that show admixture from other species (Kitajima et al., 2007; Curk et al., 2015; Garcia-Lor et al., 2015). Our sequencing of genes other than *Ruby* in satsumas (*C. unshiu*) suggested introgression from pummelo that is different from that documented in Ponkan and Willowleaf (Wu et al., 2014). *C. tachibana*, also in group A, contained a *Ruby2* allele derived from an unknown species as defined by distinct single-nucleotide polymorphisms.

Mandarins in group B include small-fruited sour mandarins (*C. sunki* and *C. reshni*) of reputed Chinese origin. Unlike $r^{DEL0.8}$, the *Ruby* allele of this group with the larger deletion (r^{DEL2}) did not show polymorphisms within or between accessions. The long 2-kb deletion encompassed the short one (0.8 kb), from which it may have been derived relatively recently.

Our data suggested that group C mandarins were derived from group B following hybridization with pummelo, an event that did not restore anthocyanin biosynthesis. This hybridization may have been a significant step in mandarin evolution that introduced new features, including fruit quality and general plant fitness.

Based on the allelic composition of *Ruby*, many mandarin varieties are likely to be hybrids between members of the three subgroups (Fig. 4). The different combinations of the three distinct alleles could help identify the origin and classify mandarin varieties that tend to cluster together in phylogenetic studies.

The allelic constitution at the *Ruby* locus also indicated the involvement of other *Citrus* species in important varieties usually classified as mandarins (Fig. 4). In particular, our analysis supports the conclusion that *C. nobilis* originated as a hybrid between mandarin (group A) and sweet orange (Hodgson, 1967). Within *C. nobilis*, the common accessions King and Kunembo inherited different *Ruby* alleles from their sweet orange parent, confirming that they had independent origins (Penjor et al., 2013; Table I; Fig. 4). Our analysis also confirmed a major contribution of papeda (*C. micrantha*) to the Nasnaran mandarin *C. amblycarpa* (Froelicher et al., 2011). *C. indica*, which was considered by both Tanaka and Swingle to be a wild mandarin species, shows no relationship to any of the other mandarins and most likely is a true species distantly related to citron (Supplemental Fig. S6).

Ruby in *Papeda*

Most *Citrus* plants are hybrids derived from different combinations of the three true species of *Citrus* (pummelo, mandarin, and citron). Contributions from a fourth species belonging to the subgenus *Papeda* have been demonstrated in some limes (Nicolosi et al., 2000). *Papeda* is considered to be an old and primitive type of *Citrus*, but its monophyletic status and its subdivision from *Citrus* are debated (Federici et al., 1998; Bayer et al., 2009; Carbonell-Caballero et al., 2015). We analyzed four different species of *Papeda* (*C. micrantha*, *C. ichangensis*, *C. macroptera*, and *C. hystrix*; Table I). All of them were able to accumulate anthocyanins in new growth but to differing extents. In *C. ichangensis*, pigmentation is strong (Fig. 1), whereas in *C. micrantha*, pigmentation is often not visible (Fig. 1) or is limited to a trace of purple on the outside of young buds (Hodgson, 1967). The *Ruby* alleles in *C. ichangensis* (R^{Pap1} and R^{Pap1}) and *C. micrantha* (R^{Pap2} and R^{Pap2}) were unrelated, while, surprisingly, *C. macroptera* and *C. hystrix* appeared to be identical hybrids of *C. micrantha* and

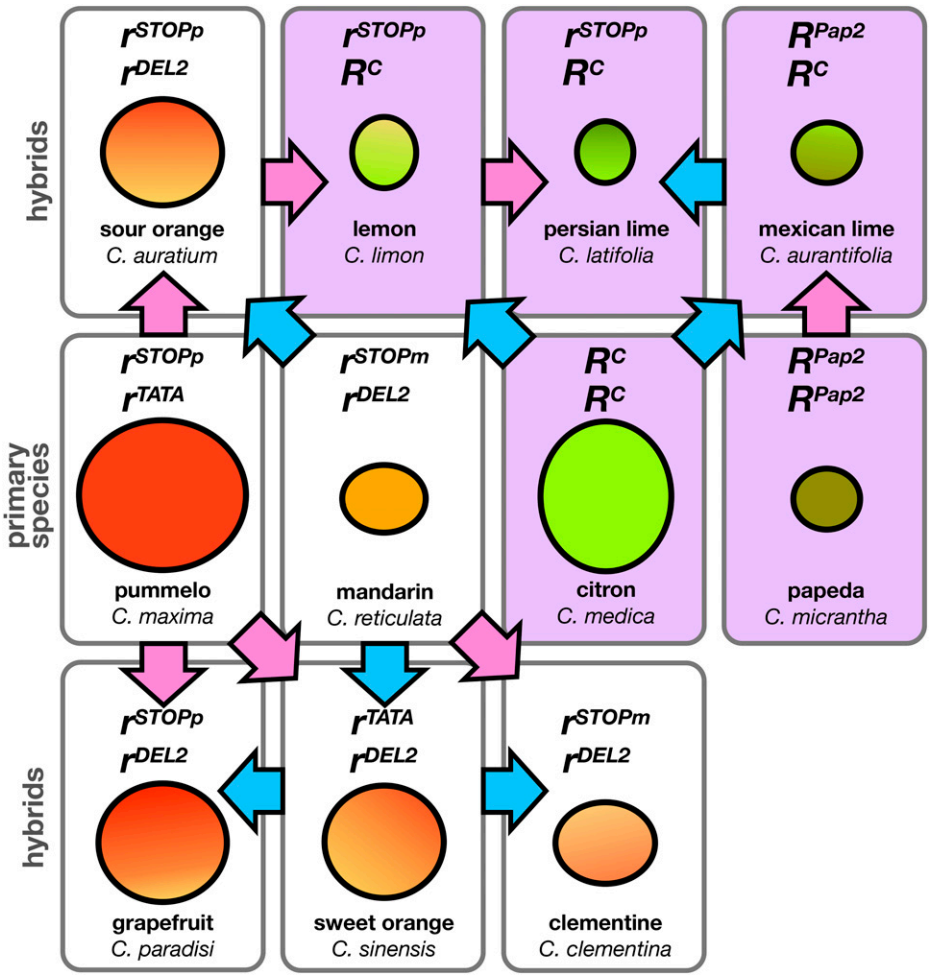

Figure 5. Segregation analysis of *Ruby* in some of the commercially important *Citrus* species and hybrids considered in this study. This simplified diagram shows a single existing variety of pummelo and mandarin as potential parents of both sweet and sour orange, which is highly unlikely. Light purple boxes indicate species and hybrids that are able to accumulate anthocyanins and that are always associated with active *Ruby* alleles from citron (R^C) or papeda (R^{Pap2}). Pink and blue arrows indicate the maternal and paternal contributions to the generation of a hybrid based on specific mitotypes and chlorotypes as determined in previous studies. The gradation in color shading indicates the contribution from different species to the derivation of the hybrids.

C. ichangensis (Table I). All the *Ruby* alleles in *Papeda* had intact open reading frames, consistent with their pigmented phenotypes. Both alleles from *C. micrantha*, however, had an insertion of a defective Copia-like retrotransposon (named Tcm7), lacking both LTRs and inserted in antisense orientation 0.3 kb upstream of the predicted start of transcription (Fig. 3). This insertion also was associated with a large deletion of the promoter region that extends to include part of the upstream *Ruby2* gene. The insertion of the retrotransposon and the deletion of part of the promoter are likely to affect the expression of *Ruby* and may explain the relatively weak pigmentation in new growth of *C. micrantha* compared with *C. ichangensis* (Fig. 1).

Ruby in Citrus Hybrids

Many commercial *Citrus* varieties have a hybrid origin (Scora, 1975; Barrett and Rhodes, 1976; Nicolosi et al., 2000; Moore, 2001). Hybrids derived from citron display intense purple pigmentation in young, developing leaves and flowers, where anthocyanins can provide protection for vulnerable tissues undergoing new growth. In contrast, hybrids involving mandarin and pummelo have inherited only nonfunctional *Ruby* alleles and do not produce anthocyanins (Fig. 5; Supplemental Fig. S7).

Citron is one parent of lemon and of several hybrids often collectively defined as limes. Citron is usually fertilized by self-pollination and always served as the male parent in the formation of cultivated hybrids (Nicolosi et al., 2000; Curk et al., 2016).

To determine the role of *Ruby* in pigmented hybrids, we focused on lemon (*C. limon*) because of the availability of a unique white mutant, Zagara Bianca, which is completely unable to accumulate anthocyanins and probably arose as a bud sport of the variety Femminello (Cultrone et al., 2010). The anthocyanin composition in young leaves of the wild type accession was determined by ultra-high-performance liquid chromatography-electrospray ionization-mass spectrometry. Two compounds, cyanidin 3-(6′′-malonyl)-β-glucoside and cyanidin 3-glucoside, accounted for most of the anthocyanin content (Fig. 6A; Supplemental Figs. S8–S10). The same compounds are the main

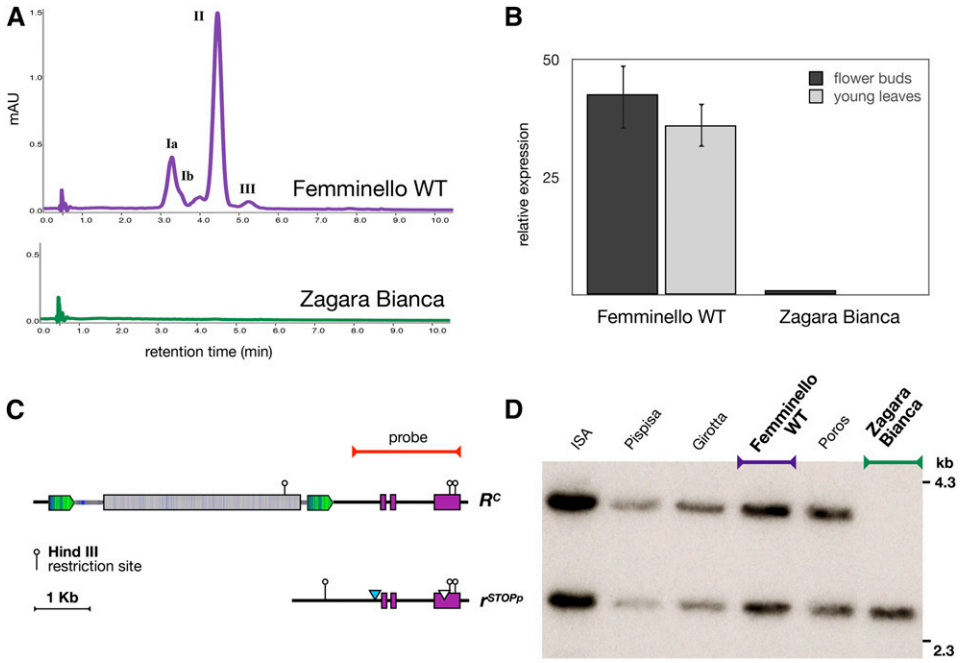

Figure 6. Analysis of the wild type (WT) and the white mutant of lemon. A, HPLC scans recorded at 525 nm of methanol extracts from young leaves of Femminello wild-type (purple line) and Zagara Bianca (green line) lemons. Peaks marked with roman numbers indicate the main compounds identified in Femminello and absent in Zagara Bianca: Ia, cyanidin 3-glucoside; Ib, cyanidin 3-rutinoside; II, cyanidin 3-(6′′-malonyl)-β-glucoside; and III, peonidin 3-(6′′-malonylglucoside). The identification of peaks is based on ultra-high-performance liquid chromatography-electrospray ionization-mass spectrometry analysis (Supplemental Figs. S8–S10) and comparison with literature data. B, Relative quantitative reverse transcription-PCR quantification of *Ruby* expression in flower buds and young leaves of Femminello and Zagara Bianca lemons. Error bars represent SE. C, Schematic diagram of the two alleles of *Ruby* identified in lemon, R^C and r^{STOPp}. The different symbols are explained in Figure 3 and Supplemental Figure S2A. The *Hin*dIII restriction sites and the genomic region used as a probe are shown. D, Southern-blot analysis of genomic DNA from different lemon varieties digested with *Hin*dIII and probed with a ^{32}P-labeled probe of the *Ruby* gene. Zagara Bianca, the only variety completely unable to produce anthocyanins, does not display the hybridization band corresponding to the active allele R^C.

anthocyanins identified in fruit of blood varieties of sweet orange and in new growth of many types of *Citrus* cultivars, including lemon (Hillebrand et al., 2004; Fabroni et al., 2016). No anthocyanins could be detected in the white mutant. *Ruby* was expressed in pigmented young flowers and leaves of the wild type accession but not in the corresponding tissues of the mutant (Fig. 6B).

The two lemon accessions were compared by Southern-blot analysis. Genomic DNA, hybridized to a probe of the full-length *Ruby* gene (Fig. 6C), displayed two distinct *Hin*dIII bands in wild-type lemon, one of which was clearly absent in the mutant (Fig. 6D). The common band present in both accessions corresponded to the *Ruby* allele r^{STOPp} carrying the nonsense mutation that disrupts protein function (Fig. 6C). This nonfunctional allele was inherited from pummelo via sour orange (Fig. 5). The allele R^C, derived from citron, was missing in the white mutant, showing that Ruby is essential for anthocyanin biosynthesis in lemon and, by inference, in other species of *Citrus*.

The active allele of *Ruby*, R^C, encodes a transcript that translates into a predicted protein of 262 amino acids with only eight differences compared with the active Ruby protein in blood varieties of sweet orange. The start of *Ruby* transcription (designated +1) in lemon was mapped using 5′ RACE. A TATA box and a potential CAAT box were identified at positions −34 and −74, respectively. The *Ruby* gene contains three exons at positions +103 to +220, +319 to +447, and +1,243 to +1,783.

Citron was involved in the independent origin of different varieties of lime. The most common are Mexican lime (*C. aurantifolia*) and Persian lime (*C. latifolia*). The composition of *Ruby* alleles in Mexican lime (R^{Pap2} and R^C) was in agreement with *Papeda* (*C. micrantha*) being its female parent (Nicolosi et al., 2000). Persian lime (r^{STOPp} and R^C) displayed the same *Ruby* alleles as lemon, which is compatible with it having been derived from a cross between lemon and Mexican lime, as proposed by Bayer et al. (2009) based on chloroplast DNA marker analysis and Curk et al. (2016) using nuclear and cytoplasmic marker analyses (Fig. 5).

The origin of other citron-derived hybrids of minor commercial importance has received little attention. Palestine sweet lime (*C. limettioides*) carried the $r^{DEL0.8}$ allele in all the accessions analyzed (including the acidic

form Soh Synteng; Table I), which argues against the contribution of either sweet or sour orange, as suggested by Nicolosi et al. (2000), and, rather, suggests the presence of a pummelo × mandarin hybrid in its parentage, as proposed by Curk et al. (2014, 2016), based on haplotyping and nuclear and cytoplasmic DNA marker analyses. This hybrid, yet to be identified, also might have been involved in the derivation of Meyer lemon, which also contains the $r^{DEL0.8}$ allele. The composition of *Ruby* alleles (r^{DEL2} and R^{C}) in limetta (*C. limetta*) was consistent with the involvement of sour orange and citron as parents (Curk et al., 2014, 2016). For Rangpur lime (*C. limonia*), the composition of *Ruby* alleles (r^{DEL2} and R^{C}) suggested a mandarin (group B) × citron origin.

The absence of anthocyanin pigmentation in flowers and leaves of many commercially important hybrids is associated with the absence of citron in their parentage (Fig. 5). For example, sour orange (*C. aurantium*), important for the production of marmalade and also used widely as a rootstock in the past, is a direct pummelo × mandarin F1 hybrid (Wu et al., 2014), confirmed by its composition of *Ruby* alleles (r^{STOPp} and r^{DEL2}), and never produces anthocyanins.

Grapefruit (*C. paradisi*) was obtained in the West Indies in the 18th century, probably as a natural hybrid between pummelo and sweet orange (Morton, 1987; Moore, 2001). This origin is supported by its *Ruby* alleles (r^{STOPp} and r^{DEL2}), and grapefruit never produces anthocyanins.

Clementine (*C. clementina*) is considered to have arisen as a natural hybrid of sweet orange and Mediterranean mandarin, based on morphological markers and sequence data (Hodgson, 1967; Nicolosi et al., 2000; Wu et al., 2014). Its lack of anthocyanins is consistent with its composition of *Ruby* alleles (r^{STOPm} and r^{DEL2}) and this hybrid origin.

Sweet orange (*C. sinensis*) is considered a complex hybrid involving mandarin (providing the predominant genetic material) and pummelo (Garcia-Lor et al., 2013b; Xu et al., 2013; Wu et al., 2014), and the allelic composition of the *Ruby* locus (r^{TATA} and r^{DEL2}) confirmed this. We have been unable to detect any expression of *Ruby* in different tissues of common varieties of sweet orange (Butelli et al., 2012), consistent with the complete absence of anthocyanin pigmentation. As discussed for pummelo, from which it is derived, the allele r^{TATA} in sweet orange is characterized by a T-to-C transition within the TATA box that prevents the gene from being expressed. In two independent blood varieties of sweet orange, however, the insertion of transposable elements has either provided a new TATA box and transcriptional start site (Tcs1) or promoted the use of a cryptic alternative start site (Tcs2) upstream of the T-to-C mutation (Supplemental Fig. S2, A and D; Butelli et al., 2012). Tcs1 and Tcs2 also provide new instructions for cold-dependent, fruit-specific expression of *Ruby* but are unable to restore the normal pattern of anthocyanin accumulation in new growth of leaves and flowers. These data confirm that r^{TATA} fails to express *Ruby* because it has a defective TATA box.

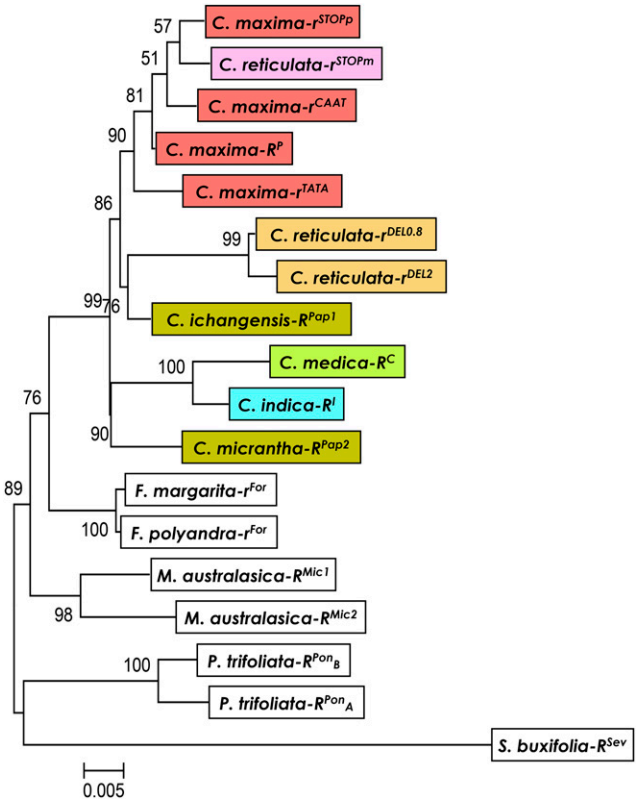

Figure 7. Phylogenetic relationship between primary species of *Citrus* and related genera based on the sequence of *Ruby*. The phylogenetic trees were generated using the Phylemon software program. Branch support is given at all branches. The different *Ruby* alleles are color coded as in Table I. The sequences of the *Ruby* gene and its promoter used for phylogenetic analysis have been deposited in GenBank under accession numbers KT591672 to KT591689.

Allelic Constitution of *Ruby* in Genera Related to *Citrus*

In addition to *Citrus*, we considered three distinct genera (*Poncirus*, *Microcitrus*, and *Fortunella*) that are classified within the true *Citrus* fruit trees and a more distant genus, *Severinia*, considered a primitive *Citrus* fruit tree (Swingle and Reece, 1967; Supplemental Fig. S1). These genera are relevant in the context of anthocyanin pigmentation.

Trifoliate orange (*Poncirus trifoliata*) is a species, unusually deciduous, that is commonly used as a rootstock for *Citrus* scions. Young and senescent *P. trifoliata* leaves are pigmented with anthocyanins (Supplemental Fig. S11A). *Ruby* is predicted to be functional in *P. trifoliata* (Fig. 3), and the two alleles we identified (R^{PonA} and R^{PonB}) were very similar in sequence, and both were expressed in pigmented leaves (Supplemental Fig. S11, B and C).

In *Microcitrus*, we initially analyzed a variety (*M. australasica* var Sanguinea) that is native to Australia, where it arose spontaneously (Netzel et al., 2006). This variety accumulates high levels of anthocyanins not only in new growth but also in fruit peel and pulp. We identified two *Ruby* alleles, R^{Mic1} and R^{Mic2}, that could be distinguished by the presence of a 0.5-kb deletion in the promoter (between positions equivalent to −887 to −364 in the R^C reference allele) of the R^{Mic1} allele or a subterminal in-frame 81-bp deletion (between positions +1,650 and +1,730) in the coding region of the R^{Mic2} allele (Fig. 3). In different tissues, the expression of *Ruby* was associated with anthocyanin pigmentation (Supplemental Fig. S12). Interestingly, while both alleles were expressed in leaves, in fruit, we could detect only monoallelic expression of R^{Mic2} (Supplemental Fig. S12E), suggesting that the subterminal deletion, affecting the amino acid sequence of a region that is not conserved in anthocyanin-specific *MYB* regulatory genes, is unlikely to affect protein function. To confirm the functionality of the transcription factor encoded by R^{Mic2}, we transiently overexpressed the corresponding cDNA in leaves of *N. benthamiana*. As shown in Figure 2C, the construct was able to induce anthocyanin biosynthesis fully with no significant differences compared with the full-length *Ruby* cDNA isolated from Moro blood orange.

Further analysis of the promoter region of R^{Mic2} revealed the presence of an insertion of a 0.4-kb repetitive element with an accompanying 7.5-kb deletion involving the entire *Ruby2* gene and a translocation of 0.5 kb of DNA from chromosome 5 at a position equivalent to −856 in the R^C reference allele (Fig. 3). The insertion, flanked by target duplication sites, corresponded to a repetitive element containing consensus motifs for SINE nonautonomous retrotransposons. Although it was difficult to assess the relative effects of the insertion, the deletion, and the translocation on the expression of the downstream *Ruby* gene, the presence of this rearrangement provided further evidence of the importance of transposable elements in reshaping gene structure and expression. This was confirmed by the analysis of a different variety of *M. australasica* with anthocyanin production in new growth but green-yellow fruit that do not accumulate anthocyanins. In this accession (Table I), the complex mutation in the promoter of *Ruby* was not present and both alleles corresponded to R^{Mic1}.

We analyzed two ornamental species belonging to the genus *Fortunella*: Nagami oval kumquat (*F. margarita*) and Malayan kumquat (*F. polyandra*), which do not accumulate anthocyanins, and we found only nonfunctional *Ruby* alleles (r^{For}) in both species, all carrying a single nucleotide insertion at position +444 in the second exon that resulted in a frame-shift mutation (Fig. 3). Calamondin (*Citrofortunella microcarpa*), another common variety that is anthocyaninless, also contained this defective allele from *Fortunella* (r^{For}) together with the $r^{DEL0.8}$ allele. This allelic composition confirmed the presence of mandarin in its ancestry but argued against a contribution by the sour variety Sunki, frequently suggested as a candidate parent (Lai and Chen, 2008).

We sequenced *Ruby* in Chinese box orange (*Severinia buxifolia*) because this species is a distant relative of the genus *Citrus* that is often used as an outgroup in phylogenetic studies. *S. buxifolia* produces highly pigmented, small fruit that becomes black at maturity. Both *Ruby* alleles were characterized by a repetitive SINE-like element of 0.5 kb inserted 0.4 kb upstream of the predicted start of transcription (Fig. 3). As in *M. australasica* var Sanguinea, it is possible that the presence of SINE-like elements could influence the strong anthocyanin pigmentation in fruit.

Phylogenetic Analysis

We conducted a phylogenetic analysis of 18 *Ruby* alleles representing the diversity of *Citrus* and related genera. Alleles from primary species only were used in the analysis, because the *Ruby* sequences in the hybrids are identical to those identified in the primary parental species (Fig. 5).

Among all of the models tested via the Phylemon Web site (Sánchez et al., 2011), the one with the best fit was HKY+G+F. The phylogenetic relationships between *Citrus* species and their relatives inferred from the maximum-likelihood method using this model are shown in Figure 7.

The model that best fitted the data using the MEGA software was the T92 model (Kimura, 1980; Tamura, 1992; Tamura et al., 2011). The phylogenetic relationships between *Citrus* species and their relatives inferred using MEGA software are shown in Supplemental Figure S13.

Phylemon and MEGA phylogenies suggested the same relationships between *Ruby* alleles, indicating a clear separation between *Citrus* and related genera (strong branch support). *Poncirus* and *Microcitrus* were quite distant from *Citrus* (high bootstrap values), while *Fortunella* was more closely related. This analysis supported the phylogenetic relationships in the Citrinae established using other markers (Nicolosi et al., 2000; Garcia-Lor et al., 2013a, 2013b; Penjor et al., 2013; Ramadugu et al., 2013) but did not reveal the distinct phylogenetic cluster of citron and *Microcitrus* established using the complete chloroplast DNA of 34 species (Carbonell-Caballero et al., 2015).

A division between the two subgenera, *Citrus* and *Papeda*, was not supported by our analysis and has been challenged in other studies (Penjor et al., 2010; Carbonell-Caballero et al., 2015). Instead, our data suggested a polyphyletic origin of the *Papeda* group. The *Ruby* alleles in *C. micrantha* are related to those of *C. medica* and *C. indica*, while the *Ruby* alleles of *C. ichangensis* clustered with the alleles from mandarin. We propose that the subgenus *Papeda* contains DNA from at least two distinct unrelated species (*C. micrantha* and *C. ichangensis*) and several hybrids between them (such as *C. hystrix* and *C. macroptera*). This distinction

could explain much of the conflicting data on the origin and classification of *Papeda*. Whole-genome sequencing would confirm or refute these suggestions.

The analysis in Figure 7 also confirmed that r^{STOPm}, one of the alleles identified in common mandarin varieties such as Ponkan and Willowleaf, lies within the clade of pummelo alleles and has been introgressed into mandarin from pummelo, as suggested by Wu et al. (2014).

DISCUSSION

Anthocyanins are widely distributed plant pigments with a diverse range of functions and tissue distributions. Since they are not essential for plant survival, they provide a unique visual marker for the occurrence of mutations and the activation of transposable elements, which may result in changes to the patterns of plant pigmentation, as demonstrated in maize (*Zea mays*), *Antirrhinum*, *Petunia*, and blood varieties of sweet orange (Malinowski, 1935; Fincham, 1967; McClintock, 1984; Coen et al., 1986; Butelli et al., 2012).

Ruby is a regulatory gene encoding a transcription factor that is essential for anthocyanin accumulation in *Citrus* and related genera. The *Ruby* locus contains two potential *MYB* regulators of anthocyanin biosynthesis (*Ruby* and the inactive *Ruby2*). Many angiosperm plants have experienced extensive duplication of these regulatory genes. The amplification of *MYB* genes belonging to subgroup 6 that regulate anthocyanin biosynthesis may be a general mechanism underlying the diversity of floral pigmentation patterns, as observed in *Petunia* (Bombarely et al., 2016) and *Antirrhinum* (Schwinn et al., 2006), but also might be the result of domestication and human selection, as proposed for grape (Matus et al., 2008). In *Citrus*, the presence of no more than one functional *MYB* regulator suggests that the production of anthocyanins is not under strong selective pressure and is largely dispensable. In Cucurbitaceae, the only family of plants where anthocyanin accumulation has never been reported, two *R2R3 MYB* genes of subgroup 6 also are present, but, according to our analysis, neither gene encodes a functional MYB activator. Therefore, while in many plant species the expansion of *MYB* genes has resulted in evolutionary novelty, in other lineages the loss of anthocyanin production is associated with molecular decay of the *MYB* regulators.

The *Ruby* locus is very dynamic. In a limited number of accessions, we uncovered six large deletions, six insertions of transposable elements, one translocation, and four critical mutations within the *Ruby* gene and its promoter region (Fig. 3). Most of these events have had consequences on anthocyanin accumulation, ranging from a complete suppression of production to complex reconfiguration of the distribution of anthocyanins in different tissues.

Transposable elements are major forces in the evolution and rearrangement of genomes and important determinants of genome size and gene expression (McClintock, 1984). The insertion of Tcl3 in citron likely did not result in obvious changes in pigmentation, but the presence of Tcm7 in *C. micrantha*, accompanied by a deletion, was associated with almost complete abolition of anthocyanin pigmentation in new growth.

The strong fruit-specific pigmentation of blood varieties of sweet orange also was the result of the insertion of retrotransposons (Butelli et al., 2012). Comparative analysis of core promoter regions in many different accessions and the mapping of the start of transcription of *Ruby* in lemon allowed us to identify a key mutation in the TATA box of *Ruby* in an allele derived from the pummelo parent in the hybrid, sweet orange. As a result, sweet orange has completely lost the ability to accumulate anthocyanins. In blood varieties of sweet orange, the insertion of retrotransposons into this nonfunctional promoter prompted the use of alternative TATA boxes that reestablished anthocyanin production and also determined the cold-induced expression of *Ruby*, specifically in fruit (Butelli et al., 2012).

The accumulation of anthocyanins in fruit of species distantly related to *Citrus* (*M. australasica* and *S. buxifolia*) also was associated with the insertion of transposable elements upstream of *Ruby*. In these species, the retrotransposons belong to a different family, SINE, and may have been involved in changing the activity of the *Ruby* promoter and the control of anthocyanin production. Normally, anthocyanin production in plants is induced by light (Steyn et al., 2002), and the evolution of anthocyanin production within the fruit might have required the development of light-independent regulatory control of *Ruby* expression, as observed with blood oranges (Butelli et al., 2012).

The loss of anthocyanins may have occurred naturally, as a consequence of the different climates in which *Citrus* species originated. Citron arose in high-altitude, dry regions where anthocyanins are important in providing protection from high light and UV irradiation (Moore, 1866; Steyn et al., 2002). In contrast, mandarin and pummelo thrive in lowland tropics with high humidity and rain throughout the year (Webber, 1943). In these conditions, light stress may be less severe and anthocyanins less important for the protection of young tissues.

Citrus plants have been grown for millennia in Asia for reasons other than for food. Citron has an ancient history as a sacred plant, and its powerful symbolism is associated with the fruit. Believed to be indigenous to northeastern India, it was absorbed into Jewish tradition and is still used in ancient religious rituals (Swingle and Reece, 1967). China is the center of origin of most *Citrus* species. In Chinese tradition, early use of *Citrus* was strictly religious and mainly associated with flowers. Kumquat flowers represented luck and prosperity, while the white flowers of mandarin and orange were symbols of purity and innocence. This symbolism has persisted into modern Christianity (Heilmeyer, 2001; Moore and Garibaldi, 2003). The simultaneous occurrence of golden fruit and white blossom remains a

central element in the symbolism of *C. tachibana*, a sacred tree in Japan. It is impossible to know the extent to which human intervention played a significant role in the selection of varieties of mandarin, pummelo, and kumquat with pure white flowers for ornamental and religious purposes. Our study, however, indicates that the ability to accumulate anthocyanins has been lost independently, several times during *Citrus* evolution, as a result of mutations in the *Ruby* gene.

Analysis of the *Ruby* locus has informed us of the origin of present-day varieties of *Citrus* and the history of their spread around the world. Although phylogenetic inference based on the analysis of a single gene may lead to conclusions that require further support, our analysis of *Ruby* has confirmed the widely accepted hybrid origin of the most important cultivated *Citrus* varieties (Fig. 5) and the phylogenetic relationships between primary species and related genera (Fig. 7). The analysis of *Ruby* also has shed new light on the origin of poorly studied hybrids, on the classification of mandarins, on members of the diverse group *Papeda*, and on the occurrence of ancient introgression events across species.

The analysis of *Ruby* explains why many *Citrus* varieties have lost the ability to accumulate anthocyanins and provides a useful tool to help in the interpretation of the mosaic structure of *Citrus* genomes.

MATERIALS AND METHODS

Plant Material

Leaves were obtained from different sources as indicated in Table I. Plant materials were ground in liquid nitrogen, and DNA and RNA were extracted using the DNeasy Plant Mini Kit and the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions.

Isolation of *Ruby* Alleles

For most accessions, *Ruby* alleles were isolated by PCR using primers designed to amplify fragments spanning the region 1.6 kb upstream of the start of the coding sequence to 136 bp downstream of the *Ruby* stop codon. PCR fragments were used directly for sequencing or cloned into the pGEM-T Easy vector (Promega). For accessions where only direct sequencing was used, nucleotide polymorphisms were depicted using the IUPAC code. The active *Ruby* alleles in citron (R^C) and in *Microcitrus australasica* var Sanguinea (R^{Mic2}) were isolated by inverse PCR after restriction enzyme digestion with *Pst*I and *Cla*I, respectively, followed by self-ligation. Primer sequences are listed in Supplemental Table S1.

DNA Gel Blot

DNA from leaves of lemons (*C. limon*) was digested with *Hin*dIII, separated by electrophoresis, transferred to nitrocellulose membranes, and hybridized as described previously (Butelli et al., 2012).

Determination of the Transcriptional Start Site of *Ruby*

The transcriptional start site in active alleles of Ruby was determined using RNA extracted from young purple leaves of Femminello lemon. DNase-treated RNA was reverse transcribed using the 5′/3′ RACE Kit (Roche). The cDNA product was sequenced to determine the transcriptional start nucleotide in the sequences lying upstream of the Ruby coding sequences. Gene-specific primers listed in Supplemental Table S1 were used for 5′ end cDNA amplification according to the manufacturer's instructions.

Expression Analysis of *Ruby*

Quantification of *Ruby* expression in new growth of Femminello and Zagara Bianca lemons and Chandler pummelo and in pigmented fruit and leaves of *M. australasica* var Sanguinea was performed by qRT-PCR as described previously (Butelli et al., 2012). The allelic expression of *Ruby* in pigmented leaves of *Poncirus trifoliata* and in fruit skin of *M. australasica* var Sanguinea was determined by RT-PCR followed by sequencing.

Transient Expression in *Nicotiana benthamiana* Leaves

Transient expression of different *Ruby* cDNA constructs was carried out using the HyperTrans system (Sainsbury et al., 2009). *Ruby* cDNAs were isolated from the flesh of Moro blood orange or the fruit peel of *M. australasica* var Sanguinea, PCR amplified using primers listed in Supplemental Table S1, and transferred into pEAQ-HT-DEST1 vectors. A truncated cDNA containing the mutations identified in the alleles r^{STOPm} and r^{STOPp} was obtained using a reverse primer with a stop codon and cDNA from Moro blood orange as a template. The plasmids obtained were introduced into *Agrobacterium tumefaciens* GV3101, and the strains were infiltrated into the abaxial side of leaves of 4-week-old plants. Leaves were harvested 7 d after infiltration.

Analysis and Identification of Anthocyanins

Leaves of lemon or *N. benthamiana* were frozen in liquid nitrogen and ground to a fine powder. Approximately 200 mg of powder was extracted with 800 μL of 80% methanol containing 1% HCl. Liquid chromatography-tandem mass spectrometry of anthocyanins was carried out on a Nexera ultra-high-performance liquid chromatography system attached to an ion-trap time of flight mass spectrometer (Shimadzu). Separation was on a 100- × 2.1-mm 2.6μ Kinetex XB-C18 column (Phenomenex) using the following gradient of acetonitrile (solvent B) versus 0.1% formic acid in water (solvent A), run at 0.5 mL min^{-1} and 40°C: 0 min, 2% B; 0.5 min, 2% B; 2 min, 10% B; 11 min, 30% B; 15 min, 90% B; 15.8 min, 90% B; 16 min, 2% B; 20.1 min, 2% B. Anthocyanins were monitored by UV/visible absorbance (200–600 nm) and positive mode electrospray mass spectrometry (full spectra from *m/z* 200 to 2,000 and data-dependent MS2 of the most abundant precursor ions at an isolation width of *m/z* 3, 50% collision energy, and 50% collision gas). Anthocyanins were identified on the basis of their mass, and the mass of their fragments in MS2, as well as by comparison with published spectra.

Phylogenetic Analysis

The sequences of the *Ruby* alleles were aligned using BioEdit (Hall, 1999), SeqMan version 7.0 (http://www.dnastar.com), and SATé-II (Liu et al., 2012) software applications. Analyses were performed to determine which model best matched the data using Phylemon 2.0 (http://phylemon.bioinfo.cipf.es; Sánchez et al., 2011). The HKY+G+F model took into account the nucleotide substitution model of Hasegawa-Kishino-Yano (HKY; number of substitution types = 2), which allowed variable base frequencies, one transition rate, and one transversion rate, as well as the nucleotide frequency (F) and the gamma distribution (G). The construction of the maximum-likelihood tree was performed using MEGA5 software with bootstrap support calculated from 1,000 samples to assess the branch support.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers KT591672 to KT591689.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Classification of *Citrus* according to Swingle.

Supplemental Figure S2. Active Ruby alleles associated with LTR retrotransposons in blood varieties of sweet orange and citron.

Supplemental Figure S3. Analysis of promoter sequences in pummelo.

Supplemental Figure S4. Multiple sequence alignment of *Ruby* in pummelo.

Supplemental Figure S5. Multiple sequence alignment of *Ruby* in mandarin.

Supplemental Figure S6. Multiple sequence alignment of *Ruby* in different species of *Citrus*.

Supplemental Figure S7. PCR analysis of *Ruby* in primary species and hybrids of *Citrus*.

Supplemental Figure S8. PCR analysis of *Ruby* in primary species and hybrids of *Citrus*.

Supplemental Figure S9. PCR analysis of *Ruby* in primary species and hybrids of *Citrus*.

Supplemental Figure S10. Mass spectrometry data for peak identification of anthocyanins in leaves of lemon.

Supplemental Figure S11. Expression analysis of *Ruby* in leaves of *P. trifoliata*.

Supplemental Figure S12. Expression analysis of *Ruby* in different tissues of *M. australasica* var Sanguinea.

Supplemental Figure S13. Phylogenetic relationship between primary species of *Citrus* and related genera based on the sequence of *Ruby* using the MEGA software program.

Supplemental Table S1. Sequences of the primers used in this study.

Supplemental Data Set S1. Nucleotide sequence of *Ruby2* in different accessions of the genus *Citrus*.

Received November 7, 2016; accepted February 10, 2017; published February 14, 2017.

LITERATURE CITED

Barrett H, Rhodes AM (1976) A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. Syst Bot **1:** 105–136

Bayer RJ, Mabberley DJ, Morton C, Miller CH, Sharma IK, Pfeil BE, Rich S, Hitchcock R, Sykes S (2009) A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. Am J Bot **96:** 668–685

Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, Bliek M, Boersma MR, Borghi L, Bruggmann R, et al (2016) Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. Nat Plants **2:** 16074

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell **24:** 1242–1255

Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J (2015) A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus Citrus. Mol Biol Evol **32:** 2015–2035

Coen ES, Carpenter R, Martin C (1986) Transposable elements generate novel spatial patterns of gene expression in Antirrhinum majus. Cell **47:** 285–296

Cooper GM, Hausman RE (2016) The Cell: A Molecular Approach. Sinauer Associates, Sunderland, MA

Cultrone A, Cotroneo PS, Recupero GR (2010) Cloning and molecular characterization of R2R3-MYB and bHLH-MYC transcription factors from Citrus sinensis. Tree Genet Genomes **6:** 101–112

Curk F, Ancillo G, Garcia-Lor A, Luro F, Perrier X, Jacquemoud-Collet JP, Navarro L, Ollitrault P (2014) Next generation haplotyping to decipher nuclear genomic interspecific admixture in Citrus species: analysis of chromosome 2. BMC Genet **15:** 152

Curk F, Ancillo G, Ollitrault F, Perrier X, Jacquemoud-Collet JP, Garcia-Lor A, Navarro L, Ollitrault P (2015) Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern citrus varieties. PLoS ONE **10:** e0125628

Curk F, Ollitrault F, Garcia-Lor A, Luro F, Navarro L, Ollitrault P (2016) Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. Ann Bot (Lond) **117:** 565–583

Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC (2007) Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. Plant J **49:** 414–427

Fabroni S, Ballistreri G, Amenta M, Rapisarda P (2016) Anthocyanins in different Citrus species: an UHPLC-PDA-ESI/MS(n)-assisted qualitative and quantitative investigation. J Sci Food Agric **96:** 4797–4808

Federici CT, Fang DQ, Scora RW, Roose ML (1998) Phylogenetic relationships within the genus Citrus (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. Theor Appl Genet **96:** 812–822

Fincham JRS (1967) Mutable genes in the light of Callan's hypothesis of serially repeated gene copies. Nature **215:** 864–866

Froelicher Y, Mouhaya W, Bassene JB, Costantino G, Kamiri M, Luro F, Morillon R, Ollitrault P (2011) New universal mitochondrial PCR markers reveal new information on maternal citrus phylogeny. Tree Genet Genomes **7:** 49–61

Garcia-Lor A, Ancillo G, Navarro L, Ollitrault P (2013a) Citrus (Rutaceae) SNP markers based on competitive allele-specific PCR: transferability across the Aurantioideae subfamily. Appl Plant Sci **1:** 1200406

Garcia-Lor A, Curk F, Snoussi-Trifa H, Morillon R, Ancillo G, Luro F, Navarro L, Ollitrault P (2013b) A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species. Ann Bot (Lond) **111:** 1–19

Garcia-Lor A, Luro F, Ollitrault P, Navarro L (2015) Genetic diversity and population structure analysis of mandarin germplasm by nuclear, chloroplastic and mitochondrial markers. Tree Genet Genomes **11:** 123

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41:** 95–98

Heilmeyer M (2001) The Language of Flowers: Symbols and Myths. Prestel USA, New York

Hillebrand S, Schwarz M, Winterhalter P (2004) Characterization of anthocyanins and pyranoanthocyanins from blood orange [*Citrus sinensis* (L.) Osbeck] juice. J Agric Food Chem **52:** 7331–7338

Hoballah ME, Gübitz T, Stuurman J, Broger L, Barone M, Mandel T, Dell'Olivo A, Arnold M, Kuhlemeier C (2007) Single gene-mediated shift in pollinator attraction in *Petunia*. Plant Cell **19:** 779–790

Hodgson RW (1967) History, world distribution, botany and varieties. *In* W Reuther, H Webber, L Batchelor, eds, The Citrus Industry. University of California Press, Berkeley, pp 431–591

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **16:** 111–120

Kitajima A, Yamasaki A, Habu T, Preedasuttijit B, Hasegawa K (2007) Chromosome identification and karyotyping of satsuma mandarin by genomic in situ hybridization. J Am Soc Hortic Sci **132:** 836–841

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science **304:** 982

Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci **10:** 236–242

Krueger R, Navarro L (2007) Citrus germplasm resources. *In* IA Khan, ed, Citrus Genetics, Breeding, and Biotechnology. CAB International, Wallingford, UK, pp 45–140

Lai YT, Chen IZ (2008) Effect of temperature on calamondin (*Citrus microcarpa*) flowering and flower bud formation. *In* Proceedings of the International Symposium on Citrus and Other Tropical and Subtropical Fruit Crops. International Society of Horticultural Science, Leuven, Belgium, pp 111–115

Liu C, Wang X, Xu Y, Deng X, Xu Q (2014) Genome-wide analysis of the R2R3-MYB transcription factor gene family in sweet orange (*Citrus sinensis*). Mol Biol Rep **41:** 6769–6785

Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu J, Stamatakis AP, Linder CR (2012) SATe-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. Syst Biol **61:** 90–106

Malinowski E (1935) Studies on unstable characters in petunia. I. The extreme flower types of the unstable race with mosaic color patterns. Genetics **20:** 342–356

Matus JT, Aquea F, Arce-Johnson P (2008) Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across Vitis and Arabidopsis genomes. BMC Plant Biol **8:** 83

McClintock B (1984) The significance of responses of the genome to challenge. Science **226:** 792–801

Moore A, Garibaldi C (2003) Flower Power: The Meaning of Flowers in Art 1500-2000. Philip Wilson Publishers, London

Moore GA (2001) Oranges and lemons: clues to the taxonomy of Citrus from molecular markers. Trends Genet **17:** 536–540

Moore T (1866) The Treasury of Botany: A Popular Dictionary of the Vegetable Kingdom, **Vol 1**. Longmans Green, London

Morton JF (1987) Grapefruit. *In* Fruits of Warm Climates. University of Florida, Miami, pp 152–158

Netzel M, Netzel G, Tian Q, Schwartz S, Konczak I (2006) Sources of antioxidant activity in Australian native fruits: identification and quantification of anthocyanins. J Agric Food Chem **54:** 9820–9826

Nicolosi E, Deng ZN, Gentile A, La Malfa S, Continella G, Tribulato E (2000) Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theor Appl Genet **100:** 1155–1166

Penjor T, Anai T, Nagano Y, Matsumoto R, Yamamoto M (2010) Phylogenetic relationships of Citrus and its relatives based on rbcL gene sequences. Tree Genet Genomes **6:** 931–939

Penjor T, Yamamoto M, Uehara M, Ide M, Matsumoto N, Matsumoto R, Nagano Y (2013) Phylogenetic relationships of citrus and its relatives based on matK gene sequences. PLoS ONE **8:** e62574

Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R (1999) Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. Plant Cell **11:** 1433–1444

Quattrocchio F, Wing JF, Leppen H, Mol J, Koes RE (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell **5:** 1497–1512

Ramadugu C, Pfeil BE, Keremane ML, Lee RF, Maureira-Butler IJ, Roose ML (2013) A six nuclear gene phylogeny of Citrus (Rutaceae) taking into account hybridization and lineage sorting. PLoS ONE **8:** e68410

Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci **10:** 63–70

Sainsbury F, Thuenemann EC, Lomonossoff GP (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol J **7:** 682–693

Sánchez R, Serra F, Tárraga J, Medina I, Carbonell J, Pulido L, de María A, Capella-Gutíerrez S, Huerta-Cepas J, Gabaldón T, et al (2011) Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Res **39:** W470–W474

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet **20:** 43–45

Schwinn K, Venail J, Shang Y, Mackay S, Alm V, Butelli E, Oyama R, Bailey P, Davies K, Martin C (2006) A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. Plant Cell **18:** 831–851

Scora RW (1975) Symposium on Biochemical Systematics, Genetics and Origin of Cultivated Plants. 9. History and Origin of Citrus. Bull Torrey Bot Club **102:** 369–375

Steyn WJ, Wand SJE, Holcroft DM, Jacobs G (2002) Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. New Phytol **155:** 349–361

Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol **4:** 447–456

Swingle WT (1946) The botany of Citrus and its wild relatives of the orange subfamily (family Rutaceae, subfamily Aurantioideae). *In* W Reuther, H Webber, L Batchelor, eds, The Citrus Industry. University of California Press, Berkeley, pp 129–474

Swingle WT, Reece PC (1967) The botany of Citrus and its wild relatives. *In* W Reuther, H Webber, L Batchelor, eds, The Citrus Industry. University of California Press, Berkeley, pp 190–430

Tamura K (1992) The rate and pattern of nucleotide substitution in Drosophila mitochondrial DNA. Mol Biol Evol **9:** 814–825

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:** 2731–2739

Tanaka T (1977) Fundamental discussion of citrus classification. Studia Citrologica **14:** 1–6

Tanaka Y, Sasaki N, Ohmiya A (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J **54:** 733–749

Webber HJ (1943) Plant characteristics and climatology. *In* W Reuther, H Webber, L Batchelor, eds, The Citrus Industry, **Vol 1**. University of California Press, Berkeley, pp 41–69

Winkel-Shirley B (2001) Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol **126:** 485–493

Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M, Scalabrin S, Terol J, et al (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat Biotechnol **32:** 656–662

Xu Q, Chen LL, Ruan X, Chen D, Zhu A, Chen C, Bertrand D, Jiao WB, Hao BH, Lyon MP, et al (2013) The draft genome of sweet orange (*Citrus sinensis*). Nat Genet **45:** 59–66