



**HAL**  
open science

## Changes in anthocyanin production during domestication of Citrus

Eugenio Butelli, Andrés Garcia-Lor, Concetta Licciardello, Giuseppina Las Casas, Lionel Hill, Giuseppe Reforgiato Recupero, Manjunath Keremane, Chandrika Ramadugu, Robert Krueger, Qiang Xu, et al.

► **To cite this version:**

Eugenio Butelli, Andrés Garcia-Lor, Concetta Licciardello, Giuseppina Las Casas, Lionel Hill, et al.. Changes in anthocyanin production during domestication of Citrus. *Plant Physiology*, 2017, 173 (4), pp.2225-2242. 10.1104/pp.16.01701 . hal-01608875

**HAL Id: hal-01608875**

**<https://hal.science/hal-01608875>**

Submitted on 29 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Document downloaded from:

[\[http://redivia.gva.es/handle/20.500.11939/6686\]](http://redivia.gva.es/handle/20.500.11939/6686)

This paper must be cited as:

[Butelli, E., Garcia-Lor, A., Licciardello, C., Las Casas, G., Hill, L., Recupero, G. R., ... & Deng, X. (2017). Changes in anthocyanin production during domestication of Citrus. *Plant Physiology*, 173(4), 2225-2242.]

**ivia**  
Institut Valencià  
d'Investigacions Agràries

The final publication is available at

[\[http://dx.doi.org/10.1104/pp.16.01701\]](http://dx.doi.org/10.1104/pp.16.01701)

Copyright [ASPB]

# Changes in anthocyanin production during domestication of *Citrus*

Eugenio Butelli<sup>1</sup>, Andrés Garcia-Lor<sup>2</sup>, Concetta Licciardello<sup>3</sup>, Giuseppina Las Casas<sup>4</sup>, Lionel Hill<sup>1</sup>, Giuseppe Reforgiato Recupero<sup>3</sup>, Manjunath L Keremane<sup>5</sup>, Chandrika Ramadugu<sup>6</sup>, Robert Krueger<sup>5</sup>, Qiang Xu<sup>7</sup>, Xiuxin Deng<sup>7</sup>, Anne-Laure Fanciullino<sup>8</sup>, Yann Froelicher<sup>9</sup>, Luis Navarro<sup>2</sup> and Cathie Martin<sup>1</sup>

<sup>1</sup>John Innes Centre, Norwich NR4 7UH, United Kingdom; <sup>2</sup>Instituto Valenciano de Investigaciones Agrarias, Carretera Moncada a Náquera Km. 4,5, 46113-Moncada, Valencia, Spain; <sup>3</sup>Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria - Centro di Ricerca per l'Agricoltura e le Colture Mediterranee, 95024 Acireale, Italy; <sup>4</sup>Dipartimento di Agricoltura, Alimentazione e Ambiente, University of Catania, 95123 Catania, Italy; <sup>5</sup>USDA-ARS National Clonal Germplasm Repository for Citrus and Dates, 1060 Martin Luther King Blvd, Riverside, CA 92507-5437, USA; <sup>6</sup>University of California Riverside, Riverside, CA 92511, USA; <sup>7</sup>Key Laboratory of Horticultural Plant Biology of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, P.R. China; <sup>8</sup>INRA, UR1115 PSH, Avignon, France; <sup>9</sup>CIRAD, UMR AGAP, Station INRA, F-20230 San Giuliano, France.

## Author contributions

E.B. and C.M. planned and designed the research.  
E.B., C.M., A.G-L., C.L., G.L.C., L.H. performed experiments.  
M.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 manuscript with input and comments from all the other authors.

## Funding Information

C.M., E.B., were supported by the Institute Strategic Program Understanding and Exploiting Plant and Microbial Secondary Metabolism (BB/J004596/1) from the BBSRC and C.M., E.B., C.L., G.L.C and G.R.R. were supported by the European Union FP7 ATHENA collaborative project (Grant Agreement 245121). AGL and LN were partially supported by grant AGL2011-26490, from the Ministry of "Economía y Competitividad" and RTA2015-00069-00-00, from the Ministry of Economía, Industria y Competitividad-INIA"- "Fondo Europeo de Desarrollo Regional (FEDER)", Spain. This work also benefited from the networking activities within the European-funded COST ACTION FA1106 QualityFruit.

43 **Author for correspondence:**

44 Eugenio Butelli: [eugenio.butelli@jic.ac.uk](mailto:eugenio.butelli@jic.ac.uk)

45 Tel: +44 (0)1603 450000

46

47 **Keywords:** anthocyanins, citrus, domestication, hybridization, MYB transcription  
48 factors, phylogenetic analysis.

49

50 **One-sentence summary**

51 *Ruby*, a regulatory gene encoding a MYB transcription factor, is essential for  
52 anthocyanin production and differences in its activity determine most of the natural  
53 variation in pigmentation in *Citrus* and related genera.

54

55

56 **Abstract**

57

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

## 76 Introduction

77

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

81 The pathway of anthocyanin biosynthesis is well characterized and the structural  
82 genes that encode the core biosynthetic enzymes have been identified from many  
83 species (Tanaka et al., 2008). These genes are coordinately induced by a conserved  
84 regulatory complex formed by the interaction between MYB transcription factors,  
85 bHLH transcription factors and WD40-repeat proteins, called the MBW complex  
86 (Quattrocchio et al., 1993; Ramsay and Glover, 2005; Koes et al., 2005). DNA-  
87 binding specificity is provided principally by the MYB factors. Mutations in the MYB  
88 regulatory genes often result in lack of expression of the biosynthetic genes and loss  
89 of anthocyanin pigmentation, as exemplified by the occurrence of 'red' and 'white'  
90 varieties of grape (Kobayashi et al., 2004), apple (Espley et al., 2007) and species of  
91 petunia (Quattrocchio et al., 1999; Hoballah et al., 2007) and antirrhinum (Schwinn et  
92 al., 2006).

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

99 The taxonomy of *Citrus*, a genus of plants belonging to the family *Rutaceae*  
100 (Supplemental Fig. S1), is complicated, as illustrated by the two most widely  
101 accepted classification systems, those proposed by Swingle and Tanaka,  
102 recognizing 16 and 162 species, respectively (Swingle, 1946; Swingle and Reece,  
103 1967; Tanaka, 1977). In *Citrus*, the definition of a species is obscured by the ability  
104 of *Citrus* species to hybridize easily within the genus and also with closely related  
105 genera. Sexual promiscuity and thousands of years of human cultivation have  
106 generated many hybrids whose origin cannot be inferred solely on the basis of  
107 morphological and geographical data. However, a growing body of evidence  
108 supports the view that there are three basic or primary species: mandarin (*C.*

109 *reticulata*), pummelo (*C. maxima*) and citron (*C. medica*) (Scora, 1975; Barrett and  
110 Rhodes, 1976). All the other 'species' are hybrids derived from these three primary  
111 species (Moore, 2001). A fourth species belonging to the subgenus *Papeda* was  
112 involved in the derivation of some limes (Nicolosi et al., 2000). The term 'secondary  
113 species' has been used widely to define *Citrus* hybrids since most of them are  
114 apomictic, polyembryonic and are normally propagated vegetatively by grafting. This  
115 ensures that plants maintain their hybrid genetic constitution over generations but  
116 also makes genetic studies challenging and crop improvement by conventional  
117 breeding, almost impossible. Within a given species or hybrid, phenotypic  
118 differences between cultivars are usually the result of spontaneous or induced  
119 somatic mutations that have been selected and propagated.

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

130

131

## 132 **Results**

133

### 134 **The *Ruby* locus in *Citrus* and related genera**

135 Most taxonomists divide the group recognized as 'true *Citrus* fruit trees' into six  
136 related genera: *Citrus*, *Poncirus*, *Microcitrus*, *Eremocitrus*, *Fortunella* and *Clymenia*  
137 (Swingle and Reece, 1967; Supplemental Fig. S1). The genus *Citrus* is further  
138 divided into two subgenera: *Citrus*, which includes the common cultivated plants with  
139 commercially important fruit, and *Papeda* (Swingle, 1946; Krueger and Navarro,  
140 2007).

141 We attempted a comprehensive analysis of the sequence variation at the *Ruby* locus

142 in the three primary *Citrus* species, in four members of the subgenus *Papeda*, in  
143 several interspecific hybrids and in accessions belonging to four genera related to  
144 *Citrus* that are relevant to understanding the regulation of anthocyanin production.  
145 For each, we sequenced the *Ruby* gene and at least 1.6 kb of the region upstream of  
146 the coding sequence. The species and hybrids we have studied are listed in Table 1.  
147 *Ruby* is involved in fruit-specific accumulation of anthocyanins in blood varieties of  
148 sweet orange (Butelli et al., 2012). A second gene, which we named *Ruby2*, also  
149 encodes a putative R2R3 MYB transcriptional activator with the potential to regulate  
150 anthocyanin biosynthesis. The *Ruby2* and *Ruby* genes are arranged in tandem on  
151 chromosome 6 and are separated by an intergenic region ranging from 1.6 to 12 kb  
152 in different species. In *Citrus*, all the *Ruby2* alleles are predicted to encode  
153 nonfunctional proteins because of the presence of different combinations of  
154 deletions, frameshift and stop mutations (Supplemental Data Set S1).  
155 No other genes with similarity to *Ruby* and encoding proteins belonging to subgroup  
156 6 of the R2R3MYB family (known to regulate anthocyanin biosynthesis; Stracke et  
157 al., 2001) were detected in the high-quality reference haploid clementine genome,  
158 nor in two different genome annotation projects of sweet orange  
159 (<https://phytozome.jgi.doe.gov>; <http://citrus.hzau.edu.cn/orange/>). A previous  
160 genome-wide targeted analysis of *R2R3 MYB* genes in sweet orange (Liu et al.,  
161 2014) and examination of the genome sequence of other *Citrus* species (QX and XD  
162 personal communication) also failed to identify other paralogs of known anthocyanin-  
163 related *MYB* genes.  
164 Based on these observations, the hypothesis was developed that, in all the  
165 accessions belonging to the genus *Citrus*, *Ruby* is the only R2R3 MYB  
166 transcriptional activator involved in anthocyanin biosynthesis. Our objective was to  
167 determine whether variation within the *Ruby* locus is responsible for the differential  
168 ability of different accessions to produce anthocyanins.

169

### 170 ***Ruby* in citron**

171 Citron (*C. medica*) is a true species with a very low level of heterozygosity. Most  
172 varieties are able to accumulate anthocyanins in young leaves and flowers (Fig. 1A),  
173 where cyanidin 3-glucoside and peonidin 3-(6"-malonyl)-glucoside accounted for  
174 approximately 80% of the total anthocyanin content (Fabroni et al., 2016).  
175 The three varieties of citron we tested (Table 1) contained the same *Ruby* allele,

176 predicted to be active and named  $R^C$ , in homozygous form.  $R^C$  was inherited,  
177 identical in sequence, in many hybrids including lemon where its expression was  
178 characterized.  
179 Lying at position -902 relative to the start of transcription in  $R^C$  is a 5435 bp  
180 retrotransposon, which we named Tc13 (Supplemental Fig. S2A and S2B). Tc13 is 92  
181 and 91% identical in sequence to the two retroelements that are responsible for  
182 anthocyanin accumulation in Mediterranean (Tcs1) and Chinese blood oranges  
183 (Tcs2) respectively (Butelli et al., 2012). All three elements are members of the same  
184 family of Copia-like LTR retrotransposons but, while Tcs1 and Tcs2 are intact, active  
185 and recently-inserted elements, Tc13 in citron is defective and is the result of an  
186 ancient retrotransposition event, as deduced by the divergence of LTR sequences  
187 that can be used to estimate the time of integration of LTR retrotransposons  
188 (SanMiguel et al., 1998). The two LTRs of Tc13 differ by seven nucleotides and a 10  
189 bp InDel, and the open reading frame of Tc13 contains four stop mutations and one 5  
190 bp deletion causing a frameshift mutation that would inactivate the encoded  
191 polyprotein (Supplemental Fig. S2B and S2C). These observations, together with its  
192 insertion at a considerable distance from the start of transcription, suggest that the  
193 presence of the Tc13 retrotransposon likely has little or no impact on *Ruby*  
194 expression.

195

### 196 ***Ruby* in pummelo**

197 Pummelo (*C. maxima*) is a true species of *Citrus* with many cultivars and high  
198 genetic diversity. Anthocyanin pigmentation has never been reported in pummelo.  
199 The four accessions we initially considered ('Chandler', 'Reinking', 'Siamese  
200 acidless' and 'Kao Pan') contain different combinations of three *Ruby* alleles that we  
201 named  $r^{STOPp}$ ,  $r^{TATA}$  and  $r^{CAAT}$  (Table 1).  
202 The  $r^{STOPp}$  allele is characterized by the presence of a stop mutation (TGG to TGA at  
203 positions equivalent to +1481 in the  $R^C$  reference allele) in the third exon that would  
204 result in the truncation of the Ruby protein involving loss of its carboxy-terminal 116  
205 amino acids. This mutation results in a complete loss of function of the Ruby protein  
206 as demonstrated by the lack of anthocyanin production when a truncated cDNA was  
207 transiently overexpressed in *Nicotiana benthamiana* (Fig. 2B). In contrast, the full-  
208 length *Ruby* cDNA induced the accumulation of visible levels of delphinidin-3-  
209 rutinoside in *N. benthamiana* (Fig. 2A).



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

224 To confirm the prediction that both  $r^{CAAT}$  and  $r^{TATA}$  alleles are not expressed, we  
225 analyzed ‘Chandler’ pummelo since both alleles are present in this variety (Table 1  
226 and Supplemental Fig. S3B). We were unable to isolate *Ruby* cDNA fragments and  
227 we could detect no *Ruby* expression in leaves of ‘Chandler’ pummelo (Supplemental  
228 Fig. S12C).

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

235 All four pummelo varieties analyzed contained at least one copy of a *Ruby* gene with  
236 an intact coding sequence. Therefore, pummelo appears to maintain the potential to  
237 produce anthocyanins, a potential that could be fulfilled by recombination of  
238 sequences lying between the mutated CAAT box allele and the mutated TATA box  
239 allele. Such an event is possible because pummelo, unlike the majority of cultivated  
240 *Citrus* plants, produces true zygotic, monoembryonic seeds. Although purple  
241 pummelo varieties have not been reported, an individual wild tree bearing purple/red-  
242 skinned fruit was identified by Prof. Xiuxin Deng in a mountainous region of Hubei  
243 province, China (Supplemental Fig. S3C). Sequence analysis indicated that this

244 pigmented variety carried a *Ruby* allele with both an 'active CAAT box' and an 'active  
245 TATA box' that might therefore be the result of such a recombination event. This  
246 unique allele, named  $R^P$ , was predicted to be fully active since an intact coding  
247 sequence was under the control of a core promoter without mutations in the CAAT or  
248 TATA boxes (Fig. 3 and Supplemental Fig. S3). Although there were mismatches in  
249 the sequence of the region upstream of the TATA box in the  $R^P$  alleles, this  
250 sequence bore a greater similarity to the equivalent regions of  $r^{TATA}$  than to  $r^{CAAT}$ ,  
251 including the absence of a 22 bp deletion. Downstream of the active TATA box motif  
252 in  $R^P$ , the sequence differed from  $r^{TATA}$  and most of the SNPs present in the area  
253 sequenced were identical between  $R^P$  and  $r^{CAAT}$ . (Supplemental Fig. S4). This new  
254 active  $R^P$  allele supports the possibility that it might be the result of recombination.  
255 Although the occurrence of such a recombination event is likely to be rare, this  
256 particular individual might have been identified and preserved by the local population  
257 because of its coloured fruit peel. The phenotype of the  $R^P$  allele might reflect the  
258 expression pattern of *Ruby* in a pigmented ancestor of pummelo, before loss-of-  
259 function *Ruby* alleles began to predominate as a result of evolution or human  
260 selection.

261

## 262 ***Ruby* in mandarin**

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

266 We analyzed two common types of mandarin: 'Ponkan', an old variety thought to be  
267 a pure, prototypical *C. reticulata* (Hodgson, 1967) and 'Avana', a Mediterranean  
268 variety also known as 'Willowleaf'. The structure and the sequence of the *Ruby* locus  
269 was identical in the two varieties and carried one allele ( $r^{STOPm}$ ) with the same stop  
270 mutation identified in pummelo and another ( $r^{DEL2}$ ) with a 2 kb deletion involving the  
271 first two exons and a large portion of the upstream regulatory region of *Ruby* (Table  
272 1, Fig. 3 and Supplemental Fig. S5). Sequence alignments suggested that the  $r^{STOPm}$   
273 allele was derived from pummelo because the two alleles,  $r^{STOPm}$  and  $r^{STOPp}$ , showed  
274 99% nucleotide identity and contained identical polymorphisms, including the  
275 distinctive stop mutation, that were not present in *Ruby* alleles from other *Citrus* or  
276 related species (Supplemental Fig. S5). Comparative sequencing of different *Citrus*  
277 varieties (Wu et al., 2014) has indicated a substantial admixture from pummelo in

278 'Ponkan' and 'Willowleaf' mandarins. In particular, the reported nucleotide  
279 heterozygosity in the terminal region of chromosome 6, where *Ruby* is located,  
280 supports the likely pummelo origin of a *Ruby* allele in some varieties of mandarin.  
281 Since an ancestral mandarin may have had a distinct allelic composition at the *Ruby*  
282 locus, we looked among existing mandarins for different alleles of *Ruby*.  
283 Tanaka recognized 36 mandarin species separated into five taxonomic groups  
284 (Tanaka, 1977). We sequenced the *Ruby* alleles from 17 mandarin varieties,  
285 representative of accessions from all five groups (Table 1).  
286 On the basis of their constitution of *Ruby* alleles, these mandarin accessions could  
287 be divided into three groups (Fig. 4). Group A ( $r^{DEL0.8}$ ,  $r^{DEL0.8}$ ) was characterized by  
288 being homozygous for a deletion of ~0.8 kb that spans the first two exons of the  
289 *Ruby* gene (between positions -175 and +629 relative to the start of transcription in  
290 the reference allele  $R^C$ ). Mandarins of group B ( $r^{DEL2}$ ,  $r^{DEL2}$ ) were homozygous for a  
291 larger deletion of ~2 kb (upstream of position +729 in  $R^C$ ). Mandarins of group C  
292 ( $r^{DEL2}$ ,  $r^{STOPm}$ ) were polymorphic for *Ruby* alleles, containing one allele introgressed  
293 from pummelo ( $r^{STOPm}$ ) and one allele with the 2 kb deletion ( $r^{DEL2}$ ) also present in  
294 group B. This polymorphic genetic constitution has probably been maintained  
295 because group C varieties, including 'Ponkan' and 'Willowleaf', are highly apomictic.  
296 None of the alleles in mandarin encode an active *Ruby* protein, explaining why  
297 mandarins have never been associated with anthocyanin pigmentation.  
298 The short deletion allele ( $r^{DEL0.8}$ ) is highly polymorphic (Supplemental Fig. S5). Group  
299 A includes varieties of Japanese origin that show admixture from other species  
300 (Kitajima, 2007; Curk et al., 2015; Garcia-Lor et al., 2015). Our sequencing of genes  
301 other than *Ruby* in satsumas (*C. unshiu*) suggested introgression from pummelo that  
302 is different from that documented in 'Ponkan' and 'Willowleaf' (Wu et al., 2014).  
303 *Citrus tachibana*, also in group A, contained a *Ruby2* allele derived from an unknown  
304 species as defined by distinct SNPs.  
305 Mandarins in group B include small-fruited sour mandarins (*C. sunki* and *C. reshni*)  
306 of reputed Chinese origin. Unlike  $r^{DEL0.8}$ , the *Ruby* allele of this group with the larger  
307 deletion ( $r^{DEL2}$ ) did not show polymorphisms within or between accessions. The long  
308 2 kb deletion encompassed the short one (0.8 kb), from which it may have been  
309 derived relatively recently.  
310 Our data suggested that group C mandarins were derived from group B following  
311 hybridization with pummelo, an event that did not restore anthocyanin biosynthesis.

312 This hybridization may have been a significant step in mandarin evolution that  
313 introduced new features including fruit quality and general plant fitness.  
314 Based on the allelic composition of *Ruby*, many mandarin varieties are likely to be  
315 hybrids between members of the three subgroups (Fig. 4). The different  
316 combinations of the three distinct alleles could help identify the origin and classify  
317 mandarin varieties that tend to cluster together in phylogenetic studies.  
318 The allelic constitution at the *Ruby* locus also indicated the involvement of other  
319 *Citrus* species in important varieties usually classified as mandarins (Fig. 4). In  
320 particular, our analysis supports the conclusion that *C. nobilis* originated as a hybrid  
321 between mandarin (group A) and sweet orange (Hodgson, 1967). Within *C. nobilis*,  
322 the common accessions 'King' and 'Kunembo' inherited different *Ruby* alleles from  
323 their sweet orange parent, confirming that they had independent origins (Penjor et  
324 al., 2013) (Table 1 and Fig. 4). Our analysis also confirmed a major contribution of  
325 papeda (*C. micrantha*) to the 'Nasranan' mandarin *C. amblycarpa* (Froelicher et al.,  
326 2011). *Citrus indica*, which was considered by both Tanaka and Swingle to be a wild  
327 mandarin species, shows no relationship to any of the other mandarins and most  
328 likely is a true species distantly related to citron (Supplemental Fig. S6).

329

### 330 ***Ruby* in *Papeda***

331 Most *Citrus* plants are hybrids derived from different combinations of the three true  
332 species of *Citrus* (pummelo, mandarin and citron). Contributions from a fourth  
333 species belonging to the subgenus *Papeda* have been demonstrated in some limes  
334 (Nicolosi et al., 2000). *Papeda* is considered to be an old and primitive type of *Citrus*  
335 but its monophyletic status and its subdivision from *Citrus* are debated (Federici et  
336 al., 1998; Bayer et al., 2009; Carbonell-Caballero et al., 2015). We analyzed four  
337 different species of *Papeda* (*C. micrantha*, *C. ichangensis*, *C. macroptera* and *C.*  
338 *hystrix*; Table 1). All of them were able to accumulate anthocyanins in new growth  
339 but to differing extents. In *C. ichangensis* pigmentation is strong (Fig. 1C) whereas in  
340 *C. micrantha* pigmentation is often not visible (Fig. 1D) or limited to a trace of purple  
341 on the outside of young buds (Hodgson, 1967). The *Ruby* alleles in *C. ichangensis*  
342 ( $R^{Pap1}$ ,  $R^{Pap1}$ ) and *C. micrantha* ( $R^{Pap2}$ ,  $R^{Pap2}$ ) were unrelated while, surprisingly, *C.*  
343 *macroptera* and *C. hystrix* appeared to be identical hybrids of *C. micrantha* and *C.*  
344 *ichangensis* (Table 1). All the *Ruby* alleles in *Papeda* had intact open reading  
345 frames, consistent with their pigmented phenotypes. Both alleles from *C. micrantha*,

346 however, had an insertion of a defective Copia-like retrotransposon (named Tcm7),  
347 lacking both LTRs and inserted in antisense orientation 0.3 kb upstream of the  
348 predicted start of transcription (Fig. 3). This insertion was also associated with a  
349 large deletion of the promoter region that extends to include part of the upstream  
350 *Ruby2* gene. The insertion of the retrotransposon and the deletion of part of the  
351 promoter are likely to affect the expression of *Ruby* and may explain the relatively  
352 weak pigmentation in new growth of *C. micrantha* compared to *C. ichangensis* (Fig.  
353 1D and 1C respectively).

354

### 355 ***Ruby* in *Citrus* hybrids**

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

363 Citron is one parent of lemon and of several hybrids often collectively defined as  
364 'limes'. Citron is usually fertilized by self-pollination and always served as the male  
365 parent in the formation of cultivated hybrids (Nicolosi et al., 2000; Curk et al., 2016).  
366 To determine the role of *Ruby* in pigmented hybrids, we focused on lemon (*C. limon*)  
367 because of the availability of a unique *white* mutant, 'Zagara Bianca', which is  
368 completely unable to accumulate anthocyanins and most probably arose as bud  
369 sport of the cultivar 'Femminello' (Cultrone et al., 2010). The anthocyanin  
370 composition in young leaves of the wild type accession was determined by UHPLC-  
371 ESI-MS. Two compounds, cyanidin 3-(6"-malonyl)- $\beta$ -glucoside and cyanidin 3-  
372 glucoside, accounted for most of the anthocyanin content (Fig. 6A and Supplemental  
373 Fig. S8, S9 and S10). The same compounds are the main anthocyanins identified in  
374 fruit of blood varieties of sweet orange and in new growth of many types of citrus  
375 cultivars, including lemon (Hillebrand et al., 2004; Fabroni et al., 2016). No  
376 anthocyanins could be detected in the *white* mutant. *Ruby*, was expressed in  
377 pigmented young flowers and leaves of the wild type accession but not in the  
378 corresponding tissues of the mutant (Fig. 6B).

379 The two lemon accessions were compared by Southern blot analysis. Genomic

380 DNA, hybridized to a probe of the full-length *Ruby* gene (Fig. 6C), displayed two  
381 distinct *Hind*III bands in wild type lemon, one of which was clearly absent in the  
382 mutant (Fig. 6D). The common band present in both accessions corresponded to the  
383 *Ruby* allele  $r^{STOPp}$  carrying the nonsense mutation that disrupts protein function (Fig.  
384 6C). This nonfunctional allele was inherited from pummelo via sour orange (Fig. 5).  
385 The allele  $R^C$ , derived from citron, was missing in the *white* mutant, showing that  
386 *Ruby* is essential for anthocyanin biosynthesis in lemon and, by inference, in other  
387 species of *Citrus*.

388 The active allele of *Ruby*,  $R^C$ , encodes a transcript which translates into a predicted  
389 protein of 262 amino acids with only eight differences compared to the active *Ruby*  
390 protein in blood varieties of sweet orange. The start of *Ruby* transcription  
391 (designated +1) in lemon was mapped using 5' RACE. A TATA box and a potential  
392 CAAT box were identified at positions -34 and -74 respectively. The *Ruby* gene  
393 contains three exons at positions +103 to +220, +319 to +447 and +1243 to +1783.  
394 Citron was involved in the independent origin of different varieties of lime. The most  
395 common are Mexican lime (*C. aurantifolia*) and Persian lime (*C. latifolia*). The  
396 composition of *Ruby* alleles in Mexican lime ( $R^{Pap2}$ ,  $R^C$ ) was in agreement with  
397 papeda (*C. micrantha*) being its female parent (Nicolosi et al., 2000). Persian lime  
398 ( $r^{STOPp}$ ,  $R^C$ ) displayed the same *Ruby* alleles as lemon, which is compatible with it  
399 having been derived from a cross between lemon and Mexican lime as proposed by  
400 Bayer et al., (2009) based on chloroplast DNA marker analysis and Curk et al.,  
401 (2016) using nuclear and cytoplasmic marker analyses (Fig. 5).

402 The origin of other citron-derived hybrids of minor commercial importance has  
403 received little attention. Palestine sweet lime (*C. limettioides*), carried the  $r^{DEL0.8}$  allele  
404 in all the accessions analyzed (including the acidic form 'Soh Synteng'; Table 1)  
405 which argues against the contribution of either sweet or sour orange as suggested  
406 by Nicolosi et al. (2000) and rather suggests the presence of a pummelo x mandarin  
407 hybrid in its parentage as proposed by Curk et al. (2014 and 2016) based on  
408 haplotyping, nuclear and cytoplasmic DNA marker analyses. This hybrid, yet to be  
409 identified, might also have been involved in the derivation of Meyer lemon, which  
410 also contains the  $r^{DEL0.8}$  allele. The composition of *Ruby* alleles ( $r^{DEL2}$ ,  $R^C$ ) in limetta  
411 (*C. limetta*) was consistent with the involvement of sour orange and citron as parents  
412 (Curk et al., 2014 and 2016). For Rangpur lime (*C. limonia*), the composition of *Ruby*  
413 alleles ( $r^{DEL2}$ ,  $R^C$ ) suggested a mandarin (group B) x citron origin.

414 The absence of anthocyanin pigmentation in flowers and leaves of many  
415 commercially important hybrids is associated with the absence of citron in their  
416 parentage (Fig. 5). For example, sour orange (*C. aurantium*), important for the  
417 production of marmalade and also used widely as a rootstock in the past, is a direct  
418 pummelo x mandarin F1 hybrid (Wu et al., 2014), confirmed by its composition of  
419 *Ruby* alleles ( $r^{STOPp}$ ,  $r^{DEL2}$ ) and never produces anthocyanins.  
420 Grapefruit (*C. paradisi*) was obtained in the West Indies in the 18<sup>th</sup> century, probably  
421 as a natural hybrid between pummelo and sweet orange (Morton, 1987; Moore,  
422 2001). This origin is supported by its *Ruby* alleles ( $r^{STOPp}$ ,  $r^{DEL2}$ ) and grapefruit never  
423 produces anthocyanins.  
424 Clementine (*C. clementina*) is considered to have arisen as a natural hybrid of sweet  
425 orange and Mediterranean mandarin based on morphological markers and sequence  
426 data (Hodgson, 1967; Nicolosi et al., 2000; Wu et al., 2014). Its lack of anthocyanins  
427 is consistent with its composition of *Ruby* alleles ( $r^{STOPm}$ ,  $r^{DEL2}$ ) and this hybrid origin.  
428 Sweet orange (*C. sinensis*) is considered a complex hybrid involving mandarin  
429 (providing the predominant genetic material) and pummelo (Garcia-Lor et al., 2013b;  
430 Xu et al., 2013; Wu et al., 2014) and the allelic composition of the *Ruby* locus ( $r^{TATA}$ ,  
431  $r^{DEL2}$ ) confirmed this. We have been unable to detect any expression of *Ruby* in  
432 different tissues of common varieties of sweet orange (Butelli et al., 2012), consistent  
433 with the complete absence of anthocyanin pigmentation. As discussed for pummelo,  
434 from which it is derived, the allele  $r^{TATA}$  in sweet orange is characterized by a T>C  
435 transition within the TATA box that prevents the gene from being expressed. In two  
436 independent blood varieties of sweet orange, however, the insertion of transposable  
437 elements has provided either a new TATA box and transcriptional start site (Tcs1), or  
438 promoted the use of a cryptic alternative start site (Tcs2) upstream of the T>C  
439 mutation (Supplemental Fig. S2A and S2D; Butelli et al., 2012). Tcs1 and Tcs2 also  
440 provide new instructions for cold-dependent, fruit-specific expression of *Ruby* but are  
441 unable to restore the normal pattern of anthocyanin accumulation in new growth of  
442 leaves and flowers. These data confirm that  $r^{TATA}$  fails to express *Ruby* because it  
443 has a defective TATA box.

444

#### 445 **Allelic constitution of *Ruby* in genera related to *Citrus***

446 In addition to *Citrus*, we considered three distinct genera - *Poncirus*, *Microcitrus* and  
447 *Fortunella* - that are classified within the 'true *Citrus* fruit trees' and a more distant

448 genus, *Severinia*, considered a 'primitive *Citrus* fruit tree' (Swingle and Reece, 1967;  
449 Supplemental Fig. S1). These genera are relevant in the context of anthocyanin  
450 pigmentation.

451 Trifoliolate orange (*Poncirus trifoliata*) is a species, unusually deciduous, that is  
452 commonly used as a rootstock for *Citrus* scions. Young and senescent *P. trifoliata*  
453 leaves are pigmented with anthocyanins (Supplemental Fig. S11A). *Ruby* is  
454 predicted to be functional in *P. trifoliata* (Fig. 3) and the two alleles we identified  
455 ( $R^{PonA}$  and  $R^{PonB}$ ) were very similar in sequence and both were expressed in  
456 pigmented leaves (Supplemental Fig. S11B and S11C).

457 In *Microcitrus*, we initially analyzed a variety (*M. australasica* var 'Sanguinea'), native  
458 to Australia, where it arose spontaneously (Netzel et al., 2007). This variety  
459 accumulates high levels of anthocyanins not only in new growth but also in fruit peel  
460 and pulp. We identified two *Ruby* alleles,  $R^{Mic1}$  and  $R^{Mic2}$ , that could be distinguished  
461 by the presence of a 0.5 kb deletion in the promoter (between positions equivalent to  
462 -887 to -364 in the  $R^C$  reference allele) of the  $R^{Mic1}$  allele or a subterminal in-frame  
463 81 bp deletion (between positions +1650 and +1730) in the coding region of the  $R^{Mic2}$   
464 allele (Fig. 3). In different tissues, the expression of *Ruby* was associated with  
465 anthocyanin pigmentation (Supplemental Fig. S12). Interestingly, while both alleles  
466 were expressed in leaves, in fruit we could detect only monoallelic expression of  
467  $R^{Mic2}$  (Supplemental Fig. S12E), suggesting that the subterminal deletion, affecting  
468 the aminoacid sequence of a region that is not conserved in anthocyanin-specific  
469 *Myb* regulatory genes, is unlikely to affect protein function. To confirm the  
470 functionality of the transcription factor encoded by  $R^{Mic2}$ , we transiently  
471 overexpressed the corresponding cDNA in leaves of *N. benthamiana*. As shown in  
472 Fig. 2C, the construct was able to induce anthocyanin biosynthesis fully with no  
473 significant differences compared to the full-length *Ruby* cDNA isolated from 'Moro'  
474 blood orange.

475 Further analysis of the promoter region of  $R^{Mic2}$  revealed the presence of an insertion  
476 of a 0.4 kb repetitive element with an accompanying 7.5 kb deletion involving the  
477 entire *Ruby2* gene and a translocation of 0.5 kb of DNA from chromosome 5 at a  
478 position equivalent to -856 in the  $R^C$  reference allele (Fig. 3). The insertion, flanked  
479 by target duplication sites, corresponded to a repetitive element containing  
480 consensus motifs for SINE nonautonomous retrotransposons. Although it was  
481 difficult to assess the relative effects of the insertion, the deletion and the



482 translocation on the expression of the downstream *Ruby* gene, the presence of this  
483 rearrangement provided further evidence of the importance of transposable elements  
484 in reshaping gene structure and expression. This was confirmed by the analysis of a  
485 different variety of *M. australasica* with anthocyanin production in new growth but  
486 green-yellow fruit that do not accumulate anthocyanins. In this accession (Table 1),  
487 the complex mutation in the promoter of *Ruby* was not present and both alleles  
488 corresponded to  $R^{Mic1}$ .

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

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

507

### 508 **Phylogenetic analysis**

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

513 Among all of the models tested via the Phylemon website (Sánchez et al., 2011), the  
514 one with the best fit was HKY+G+F. The phylogenetic relationships between *Citrus*  
515 species and their relatives inferred from the maximum-likelihood (ML) method using

516 this model are shown in Fig. 7.  
517 The model that best fitted the data using the MEGA (Molecular Evolutionary  
518 Genetics Analysis) software was the T92 model (Kimura, 1980; Tamura, 1992;  
519 Tamura et al., 2011). The phylogenetic relationships between *Citrus* species and  
520 their relatives inferred using MEGA software are shown in Supplemental Fig. S13.  
521 Phylemon and MEGA phylogenies suggested the same relationships between *Ruby*  
522 alleles indicating a clear separation between *Citrus* and related genera (strong  
523 branch support). *Poncirus* and *Microcitrus* were quite distant from *Citrus* (high  
524 bootstrap values) while *Fortunella* was more closely related. This analysis supported  
525 the phylogenetic relationships in the *Citrinae* established using other markers  
526 (Nicolosi et al., 2000; Garcia-Lor et al., 2013a and 2013b; Ramadugu et al., 2013;  
527 Penjor et al., 2013) but did not reveal the distinct phylogenetic cluster of citron and  
528 *Microcitrus* established using the complete chloroplast DNA of 34 species  
529 (Carbonell-Caballero et al., 2015).  
530 A division between the two subgenera, *Citrus* and *Papeda*, was not supported by our  
531 analysis and has been challenged in other studies (Penjor et al., 2010; Carbonell-  
532 Caballero et al., 2015). Instead our data suggested a polyphyletic origin of the  
533 *Papeda* group. The *Ruby* alleles in *C. micrantha* are related to those of *C. medica*  
534 and *C. indica* while the *Ruby* alleles of *C. ichangensis* clustered with the alleles from  
535 mandarin. We propose that the subgenus *Papeda* contains DNA from at least two  
536 distinct unrelated species (*C. micrantha* and *C. ichangensis*) and several hybrids  
537 between them (such as *C. hystrix* and *C. macroptera*). This distinction could explain  
538 much of the conflicting data on the origin and classification of *Papeda*. Whole  
539 genome sequencing would confirm or refute these suggestions.  
540 The analysis in Fig. 7 also confirmed that  $r^{STOPm}$ , one of the alleles identified in  
541 common mandarin varieties such as 'Ponkan' and 'Willowleaf', lies within the clade  
542 of pummelo alleles and has been introgressed into mandarin from pummelo as  
543 suggested by Wu et al. (2014).

544  
545

## 546 **Discussion**

547

548 Anthocyanins are widely distributed plant pigments with a diverse range of functions

549 and tissue distributions. Since they are not essential for plant survival, they provide a  
550 unique visual marker for the occurrence of mutations and activation of transposable  
551 elements, which may result in changes to the patterns of plant pigmentation as  
552 demonstrated in maize, antirrhinum, petunia and blood varieties of sweet orange  
553 (Malinowski, 1935; Fincham, 1967; McClintock, 1984; Coen et al., 1986; Butelli et al.,  
554 2012).

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

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

578 Transposable elements are major forces in the evolution and rearrangement of  
579 genomes and important determinants of genome size and gene expression  
580 (McClintock, 1984). The insertion of Tc13 in citron likely did not result in obvious  
581 changes in pigmentation but the presence of Tcm7 in *C. micrantha*, accompanied by  
582 a deletion, was associated with almost complete abolition of anthocyanin

583 pigmentation in new growth.

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

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

603 The loss of anthocyanins may have occurred naturally, as a consequence of the  
604 different climates in which *Citrus* species originated. Citron arose in high-altitude, dry  
605 regions where anthocyanins are important in providing protection from high light and  
606 UV irradiation (Moore, 1866; Steyn et al., 2002). In contrast, mandarin and pummelo  
607 thrive in lowland tropics with high humidity and rain throughout the year (Webber,  
608 1943). In these conditions, light stress may be less severe and anthocyanins less  
609 important for protection of young tissues.

610 *Citrus* plants have been grown for millennia in Asia for reasons other than for food.  
611 Citron has an ancient history as a sacred plant and its powerful symbolism is  
612 associated with the fruit. Believed to be indigenous to Northeastern India, it was  
613 absorbed into Jewish tradition and is still used in ancient religious rituals (Swingle  
614 and Reece, 1967). China is the centre of origin of most citrus species. In Chinese  
615 tradition, early use of *Citrus* was strictly religious and mainly associated with flowers.  
616 Kumquat flowers represented luck and prosperity while the white flowers of mandarin

617 and orange were symbols of purity and innocence. This symbolism has persisted  
618 into modern Christianity (Heilmeyer, 2001; Moore and Garibaldi, 2003). The  
619 simultaneous occurrence of golden fruit and white blossom remains a central  
620 element in the symbolism of *C. tachibana*, a sacred tree in Japan. It is impossible to  
621 know the extent to which human intervention played a significant role in the selection  
622 of varieties of mandarin, pummelo and kumquat with pure white flowers for  
623 ornamental and religious purposes. Our study, however, indicates that the ability to  
624 accumulate anthocyanins has been lost independently, several times during *Citrus*  
625 evolution, as a result of mutations in the *Ruby* gene.

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

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

638

639

## 640 **Materials and Methods**

641

### 642 **Plant Material**

643 Leaves were obtained from different sources as indicated in Table 1. Plant materials  
644 were ground in liquid nitrogen and DNA and RNA were extracted using DNeasy  
645 Plant Mini Kit and RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's  
646 instructions.

647

### 648 **Isolation of *Ruby* alleles**

649 For most accessions, *Ruby* alleles were isolated by PCR using primers designed to

650 amplify fragments spanning the region 1.6 kp upstream of the start of the coding  
651 sequence to 136 bp downstream of the *Ruby* stop codon. PCR fragments were used  
652 directly for sequencing or cloned into the pGEM-T Easy vector (Promega). For  
653 accessions where only direct sequencing was used, nucleotide polymorphisms were  
654 depicted using the IUPAC code. The active *Ruby* alleles in citron ( $R^C$ ) and in *M.*  
655 *australasica* var 'Sanguinea' ( $R^{Mic2}$ ) were isolated by inverse PCR after restriction  
656 enzyme digestion with *Pst*I and *Cla*I respectively followed by self-ligation. Primer  
657 sequences are listed in Supplemental Table S1.

658

### 659 **DNA Gel Blot**

660 DNA from leaves of lemons was digested with *Hind*III, separated by electrophoresis,  
661 transferred to nitrocellulose membranes and hybridized as described previously  
662 (Butelli et al., 2012).

663

### 664 **Determination of the transcriptional start site of *Ruby***

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

672

### 673 **Expression analysis of *Ruby***

674 Quantification of *Ruby* expression in new growth of 'Femminello' and 'Zagara Bianca'  
675 lemons, 'Chandler' pummelo and in pigmented fruit and leaves of *M. australasica* var  
676 'Sanguinea' was performed by qRT-PCR as previously described (Butelli et al.,  
677 2012). The allelic expression of *Ruby* in pigmented leaves of *P. trifoliata* and in fruit  
678 skin of *M. australasica* var 'Sanguinea' was determined by RT-PCR followed by  
679 sequencing.

680

### 681 **Transient expression in *N. benthamiana* leaves**

682 Transient expression of different *Ruby* cDNA constructs was carried out using the  
683 HyperTrans system (Sainsbury et al., 2009). *Ruby* cDNAs were isolated from the

684 flesh of 'Moro' blood orange or the fruit peel of *M. australasica* var 'Sanguinea', PCR  
685 amplified using primers listed in Supplemental Table S1 and transferred into pEAQ-  
686 HT-DEST1 vectors. A truncated cDNA containing the mutations identified in the  
687 alleles  $r^{STOPm}$  and  $r^{STOPp}$  was obtained using a reverse primer with a stop codon and  
688 cDNA from 'Moro' blood orange as template. The plasmids obtained were introduced  
689 into *Agrobacterium tumefaciens* GV3101 and the strains were infiltrated into the  
690 abaxial side of leaves of four-week-old plants. Leaves were harvested 7 days after  
691 infiltration.

692

### 693 **Analysis and identification of anthocyanins**

694 Leaves of lemon or *N. benthamiana* were frozen in liquid nitrogen and ground to a  
695 fine powder. Approximately 200 mg of powder was extracted with 800  $\mu$ L of 80%  
696 methanol containing 1% HCl. LC MS/MS of anthocyanins was carried out on a  
697 Nexera UHPLC system attached to an Ion-trap ToF (IT-ToF) mass spectrometer  
698 (Shimadzu). Separation was on a 100 $\times$ 2.1mm 2.6 $\mu$  Kinetex XB-C18 column  
699 (Phenomenex) using the following gradient of acetonitrile (solvent B) versus 0.1%  
700 formic acid in water (solvent A), run at 0.5mL.min<sup>-1</sup> and 40°C: 0 min, 2% B; 0.5 min,  
701 2% B; 2 min, 10% B; 11 min, 30% B; 15 min, 90% B; 15.8 min, 90% B; 16 min, 2%  
702 B; 20.1 min, 2% B. Anthocyanins were monitored by UV/visible absorbance (200-  
703 600nm) and positive mode electrospray MS (full spectra from  $m/z$  200-2000 and  
704 data-dependent MS2 of the most abundant precursor ions at an isolation width of  
705  $m/z$  3.0, 50% collision energy and 50% collision gas). Anthocyanins were identified  
706 on the basis of their mass, and the mass of their fragments in MS2 as well as by  
707 comparison to published spectra.

708

### 709 **Phylogenetic analysis**

710 The sequences of the *Ruby* alleles were aligned using BioEdit (Hall, 1999), SeqMan  
711 version 7.0 (<http://www.dnastar.com>) and SATé-II (Liu et al., 2012) software  
712 applications. Analyses were performed to determine which model best matched the  
713 data using Phylemon 2.0 (<http://phylemon.bioinfo.cipf.es>; Sánchez et al., 2011). The  
714 HKY+G+F model took into account the nucleotide substitution model of Hasegawa-  
715 Kishino-Yano (HKY,  $nst=2$ ), which allowed variable base frequencies, one transition  
716 rate and one transversion rate, and took into account the nucleotide frequency (F)  
717 and the gamma distribution (G). The construction of the maximum-likelihood (ML)

718 tree was performed using MEGA5 software with bootstrap support calculated from  
719 1000 samples to assess the branch support.

720

721

## 722 **Supplemental Data**

723

724 The following supplemental materials are available.

725

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

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

729 **Supplemental Figure S3.** Analysis of promoter sequences in pummelo.

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

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

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

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

734 **Supplemental Figure S8, S9 and S10.** Mass spectrometry data for peak identification of  
735 anthocyanins in leaves of lemon.

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

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

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

741 **Supplemental Table S1.** Sequences of the primers used in the study.

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

744

745

## 746 **Figure and Table Legends**

747

748 **Table 1** *Citrus* accessions were obtained from the following sources: **A:** IVIA, Instituto Valenciano de  
749 Investigaciones Agrarias, Moncada, Valencia, Spain; **B:** CREA-CRA, Centro di Ricerca per  
750 l'Agromicoltura e le Colture Mediterranee, Acireale, Italy; **C:** USDA-ARS National Clonal Germplasm  
751 Repository for Citrus & Dates, Riverside, CA, USA; **D:** CRB CITRUS, INRA-CIRAD, Citrus Biological  
752 Resource Center, San Giuliano, Corsica, France; **E:** Huazhong Agricultural University, Wuhan, China;  
753 **F:** Jingzhou, Huannan Province, China; **G:** Waitrose Supermarket, Norwich, UK; **H:** Notcutts Garden  
754 Centre, Norwich, UK.

755 In accessions of the genus *Citrus*, the *Ruby* alleles are colour-coded as follows: yellow, alleles

756 derived from the mandarin pool; pink, alleles in mandarin introgressed from pummelo; red, alleles



757 derived from pummelo; green, alleles derived from citron; olive green, alleles derived from *Papeda*;  
758 the unique allele in *C. indica* is in blue; the alleles in genera related to *Citrus* are not coloured. The  
759 lemon accession indicated with an asterisk (\*) is a 'graft chimera' and contains a third *Ruby* allele that  
760 may have originated from an acidless limetta. Pictures of fruit and flowers of the accessions indicated  
761 in bold are provided in Fig.1.

762

763

764 **Figure 1.** Phenotypes of some of the accessions considered in this study and listed in Table1. For  
765 each accession, the common and the scientific name according to Tanka's classification are provided,  
766 together with the allelic constitution at the *Ruby* locus. Images of fruit and flowers or young leaves are  
767 presented to show the presence or absence of anthocyanin pigmentation. Scale bar is approximately  
768 1 cm. The images were obtained from the following sources: ◆, pictures taken by the authors; ☞,  
769 UC-Riverside Citrus Variety Collection (<http://www.citrusvariety.ucr.edu/index.html>); ☞, Citrus ID  
770 Tools (<http://idtools.org/id/citrus/citrusid/>); ⚡, Radoslav Krchnavek; ♂, Eve and George DeLange.

771

772 **Figure 2.** Functionality of different *Ruby* alleles. *Nicotiana benthamiana* leaves were infiltrated with  
773 cDNA constructs corresponding to three different *Ruby* alleles. The phenotypes of leaves, the colour  
774 of the leaf extracts and the HPLC chromatograms at 525 nm 7 days after infiltration are shown. A,  
775 Transient overexpression of *Ruby* cDNA encoding a full-length protein isolated from the flesh of  
776 'Moro' blood orange. B, Transient overexpression of a cDNA construct encoding a protein with a  
777 terminal deletion of 116 amino acids corresponding to the alleles found in some pummelo and  
778 mandarin accessions and hybrids derived from them. C, Transient overexpression of *Ruby* cDNA  
779 isolated from the fruit peel of *Microcitrus australasica* var 'Sanguinea' encoding a protein with a 27  
780 amino acid subterminal deletion. D, Evidence of the identity of delphinidin-3-rutinoside in sample C by  
781 positive mode electrospray MS1 and MS2. For this compound the expected mass is 611.1607 (found  
782 mass 611.1633; error 4.25ppm). Note the fragment representing delphinidin (expected mass  
783 303.0499) and the losses of a hexose moiety (expected loss 162.0528) and rhamnose moiety  
784 (expected loss 146.0579).

785

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

789

790 **Figure 4.** Classification of mandarins based on the analysis of the *Ruby* gene. Mandarins can be  
791 classified into three groups (A, B and C) on the basis of two deletions and an introgression from  
792 pummelo. Many mandarin varieties are hybrids between members of these three groups or  
793 interspecific hybrids involving other *Citrus* species. *Citrus indica*, recognized by both Tanaka and  
794 Swingle as a mandarin species, is not related to mandarins. Roman numbers indicate the five  
795 taxonomic groups recognized by Tanaka (Hodgson, 1967). The gradation in colour shading indicates  
796 the contribution from genomes of other species. Common names for the mandarin accessions are  
797 provided in Table1.

798

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

807

808 **Figure 6.** Analysis of wild type and *white* mutant of lemon. A, HPLC chromatogram recorded at 525  
809 nm of methanol extracts from young leaves of 'Femminello' wild type (purple line) and 'Zagara Bianca'  
810 (green line) lemons. Peaks marked with roman numbers indicate the main compounds identified in  
811 'Femminello' and absent in 'Zagara Bianca'. **Ia**, cyanidin 3-glucoside; **Ib**, cyanidin 3-rutinoside; **II**,  
812 cyanidin 3-(6"-malonyl)- $\beta$ -glucoside; **III**, peonidin 3-(6"- malonylglucoside). Identification of peaks is  
813 based on UHPLC-ESI-MS analysis (Supplemental Fig. S8, S9 and S10) and comparison to literature  
814 data. B, Relative qRT-PCR quantification of *Ruby* expression in flower buds and young leaves of  
815 'Femminello' and 'Zagara Bianca' lemons. Error bars represent SE of the mean. C, Schematic  
816 diagram of the two alleles of *Ruby* identified in lemon,  $R^C$  and  $r^{STOPp}$ . The different symbols are  
817 explained in Fig. 3 and Supplemental Fig. S2A. The *HindIII* restriction sites and the genomic region  
818 used as a probe are shown. D, Southern blot analysis of genomic DNA from different lemon varieties  
819 digested with *HindIII* and probed with a  $^{32}P$ -labelled probe of the *Ruby* gene. 'Zagara Bianca', the only  
820 variety completely unable to produce anthocyanins, doesn't display the hybridization band  
821 corresponding to the active allele  $R^C$ .

822

823 **Figure 7.** Phylogenetic relationship between primary species of *Citrus* and related genera based on  
824 the sequence of *Ruby*. The phylogenetic trees were generated using Phylemon software program.  
825 Branch support is given in all branches. The different *Ruby* alleles are colour-coded as in Table 1.  
826 The sequences of the *Ruby* gene and its promoter used for phylogenetic analysis have been  
827 deposited in GenBank under accession numbers KT591672-KT591689.

828

829

## 830 References

831

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

834

835 **Bayer RJ, Mabberley DJ, Morton C, Miller CH, Sharma IK, Pfeil BE, Rich S,**  
836 **Hitchcock R, Sykes S** (2009) A Molecular Phylogeny of the Orange Subfamily

837 (Rutaceae: Aurantioideae) Using Nine Cpdna Sequences. American Journal of  
838 Botany **96**: 668-685  
839

840 **Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, Bliet M,**  
841 **Boersma MR, Borghi L, Bruggmann R et al.** (2016) Insight into the evolution  
842 of the Solanaceae from the parental genomes of *Petunia hybrida*. Nature Plants  
843 **2**: 16074  
844

845 **Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-**  
846 **Recupero G, Martin C** (2012) Retrotransposons control fruit-specific, cold-  
847 dependent accumulation of anthocyanins in blood oranges. Plant Cell **24**: 1242-  
848 1255  
849

850 **Carbonell-Caballero J, Alonso R, Ibanez V, Terol J, Talon M, Dopazo J** (2015) A  
851 Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships  
852 between Wild and Domestic Species within the Genus *Citrus*. Molecular Biology  
853 and Evolution **32**: 2015-2035  
854

855 **Coen ES, Carpenter R, Martin C** (1986) Transposable Elements Generate Novel  
856 Spatial Patterns of Gene-Expression in *Antirrhinum Majus*. Cell **47**: 285-296  
857

858 **Cooper GM, Hausman RE** (2016) The Cell: A Molecular Approach. Sinauer  
859 Associates, Sunderland (MA)  
860

861 **Cultrone A, Cotroneo PS, Recupero GR** (2010) Cloning and molecular  
862 characterization of R2R3-MYB and bHLH-MYC transcription factors from *Citrus*  
863 *sinensis*. Tree Genetics & Genomes **6**: 101-112  
864

865 **Curk F, Ancillo G, Garcia-Lor A, Luro F, Perrier X, Jacquemoud-Collet JP,**  
866 **Navarro L, Ollitrault P** (2014) Next generation haplotyping to decipher nuclear  
867 genomic interspecific admixture in *Citrus* species: analysis of chromosome 2.  
868 Bmc Genetics **15**: 152  
869

870 **Curk F, Ancillo G, Ollitrault F, Perrier X, Jacquemoud-Collet JP, Garcia-Lor A,**  
871 **Navarro L, Ollitrault P** (2015) Nuclear Species-Diagnostic SNP Markers Mined  
872 from 454 Amplicon Sequencing Reveal Admixture Genomic Structure of  
873 Modern *Citrus* Varieties. Plos One **10**: e0125628  
874

875 **Curk F, Ollitrault F, Garcia-Lor A, Luro F, Navarro L, Ollitrault P** (2016)  
876 Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear  
877 markers. Annals of Botany **117**: 565-583  
878

879 **Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC**  
880 (2007) Red colouration in apple fruit is due to the activity of the MYB  
881 transcription factor, MdMYB10. Plant Journal **49**: 414-427  
882

883 **Fabroni S, Ballistreri G, Amenta M, Rapisarda P** (2016) Anthocyanins in different  
884 *Citrus* species: an UHPLC-PDA-ESI/MSn-assisted qualitative and quantitative  
885 investigation. Journal of the Science of Food and Agriculture **96**: 4797-4808  
886

- 887 **Federici CT, Fang DQ, Scora RW, Roose ML** (1998) Phylogenetic relationships  
888 within the genus *Citrus* (Rutaceae) and related genera as revealed by RFLP  
889 and RAPD analysis. *Theoretical and Applied Genetics* **96**: 812-822  
890
- 891 **Fincham JRS** (1967) Mutable Genes in Light of Callans Hypothesis of Serially  
892 Repeated Gene Copies. *Nature* **215**: 864-866  
893
- 894 **Froelicher Y, Mouhaya W, Bassene JB, Costantino G, Kamiri M, Luro F,**  
895 **Morillon R, Ollitrault P** (2011) New universal mitochondrial PCR markers  
896 reveal new information on maternal citrus phylogeny. *Tree Genetics &*  
897 *Genomes* **7**: 49-61  
898
- 899 **Garcia-Lor A, Ancillo G, Navarro L, Ollitrault P** (2013) Citrus (Rutaceae) SNP  
900 Markers Based on Competitive Allele-Specific PCR; Transferability across the  
901 Aurantioideae Subfamily. *Applications in Plant Sciences* **1**: 1200406  
902
- 903 **Garcia-Lor A, Curk F, Snoussi-Trifa H, Morillon R, Ancillo G, Luro F, Navarro L,**  
904 **Ollitrault P** (2013) A nuclear phylogenetic analysis: SNPs, indels and SSRs  
905 deliver new insights into the relationships in the 'true citrus fruit trees' group  
906 (*Citrinae*, Rutaceae) and the origin of cultivated species. *Annals of Botany* **111**:  
907 1-19  
908
- 909 **Garcia-Lor A, Luro F, Ollitrault P, Navarro L** (2015) Genetic diversity and  
910 population structure analysis of mandarin germplasm by nuclear, chloroplastic  
911 and mitochondrial markers. *Tree Genetics & Genomes* **11**: 123  
912
- 913 **Hall T** (1999) BioEdit: a user-friendly biological sequence alignment editor and  
914 analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95-98  
915
- 916 **Heilmeyer M** (2001) *The Language of Flowers: Symbols and Myths*. Prestel USA,  
917 New York  
918
- 919 **Hillebrand S, Schwarz M, Winterhalter P** (2004) Characterization of anthocyanins  
920 and pyranoanthocyanins from blood orange [*Citrus sinensis* (L.) Osbeck] juice.  
921 *Journal of Agricultural and Food Chemistry* **52**: 7331-7338  
922
- 923 **Hoballah ME, Gubitz T, Stuurman J, Broger L, Barone M, Mandel T, Dell'Olivo**  
924 **A, Arnold M, Kuhlemeier C** (2007) Single gene-mediated shift in pollinator  
925 attraction in *Petunia*. *Plant Cell* **19**: 779-790  
926
- 927 **Hodgson RW** (1967) History, world distribution, botany and varieties. *In* W Reuther,  
928 H Webber, L Batchelor, eds, *The Citrus Industry*. University of California Press,  
929 Berkeley, CA, pp 431-591  
930
- 931 **Kimura M** (1980) A Simple Method for Estimating Evolutionary Rates of Base  
932 Substitutions through Comparative Studies of Nucleotide-Sequences. *Journal*  
933 *of Molecular Evolution* **16**: 111-120  
934
- 935 **Kitajima A, Yamasaki A, Habu T, Preedasuttijit B, Hasegawa K** (2007)  
936 Chromosome identification and karyotyping of satsuma mandarin by genomic in

- 937 situ hybridization. *Journal of the American Society for Horticultural Science* **132**:  
938 836-841  
939
- 940 **Kobayashi S, Goto-Yamamoto N, Hirochika H** (2004) Retrotransposon-induced  
941 mutations in grape skin color. *Science* **304**: 982  
942
- 943 **Koes R, Verweij W, Quattrocchio F** (2005) Flavonoids: a colorful model for the  
944 regulation and evolution of biochemical pathways. *Trends in Plant Science* **10**:  
945 236-242  
946
- 947 **Krueger R, Navarro L** (2007) Citrus germplasm resources. *In* IA Khan, ed, Citrus  
948 genetics, breeding, and biotechnology. CAB International, Wallingford, UK, pp  
949 45-140  
950
- 951 **Lai YT, Chen IZ** (2008) Effect of Temperature on Calamondin (*Citrus microcarpa*)  
952 Flowering and Flower Bud Formation. *Proceedings of the International*  
953 *Symposium on Citrus and Other Tropical and Subtropical Fruit Crops*: 111-115  
954
- 955 **Liu CY, Wang X, Xu YT, Deng XX, Xu Q** (2014) Genome-wide analysis of the  
956 R2R3-MYB transcription factor gene family in sweet orange (*Citrus sinensis*).  
957 *Molecular Biology Reports* **41**: 6769-6785  
958
- 959 **Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu JY, Stamatakis AP, Linder CR**  
960 (2012) SATE-II: Very Fast and Accurate Simultaneous Estimation of Multiple  
961 Sequence Alignments and Phylogenetic Trees. *Systematic Biology* **61**: 90-106  
962
- 963 **Malinowski E** (1935) Studies on Unstable characters in *Petunia*. The extreme flower  
964 types of the Unstable race with mosaic color patterns. *Genetics* **20**: 342-356  
965
- 966 **Matus JT, Aquea F, Arce-Johnson P** (2008) Analysis of the grape MYB R2R3  
967 subfamily reveals expanded wine quality-related clades and conserved gene  
968 structure organization across *Vitis* and *Arabidopsis* genomes. *Bmc Plant*  
969 *Biology* **8**: 83  
970
- 971 **McClintock B** (1984) The Significance of Responses of the Genome to Challenge.  
972 *Science* **226**: 792-801  
973
- 974 **Moore A, Garibaldi C** (2003) Flower Power: The Meaning of Flowers in Art 1500-  
975 2000. Philip Wilson Publishers, London  
976
- 977 **Moore GA** (2001) Oranges and lemons: clues to the taxonomy of Citrus from  
978 molecular markers. *Trends in Genetics* **17**: 536-540  
979
- 980 **Moore T** (1866 ) *The Treasury of Botany: a popular dictionary of the vegetable*  
981 *kingdom. , Vol 1.* Longmans Green & Co., London  
982
- 983 **Morton JF** (1987) Grapefruit. *In* *Fruits of warm climates.* University of Florida, Miami,  
984 pp 152-158  
985

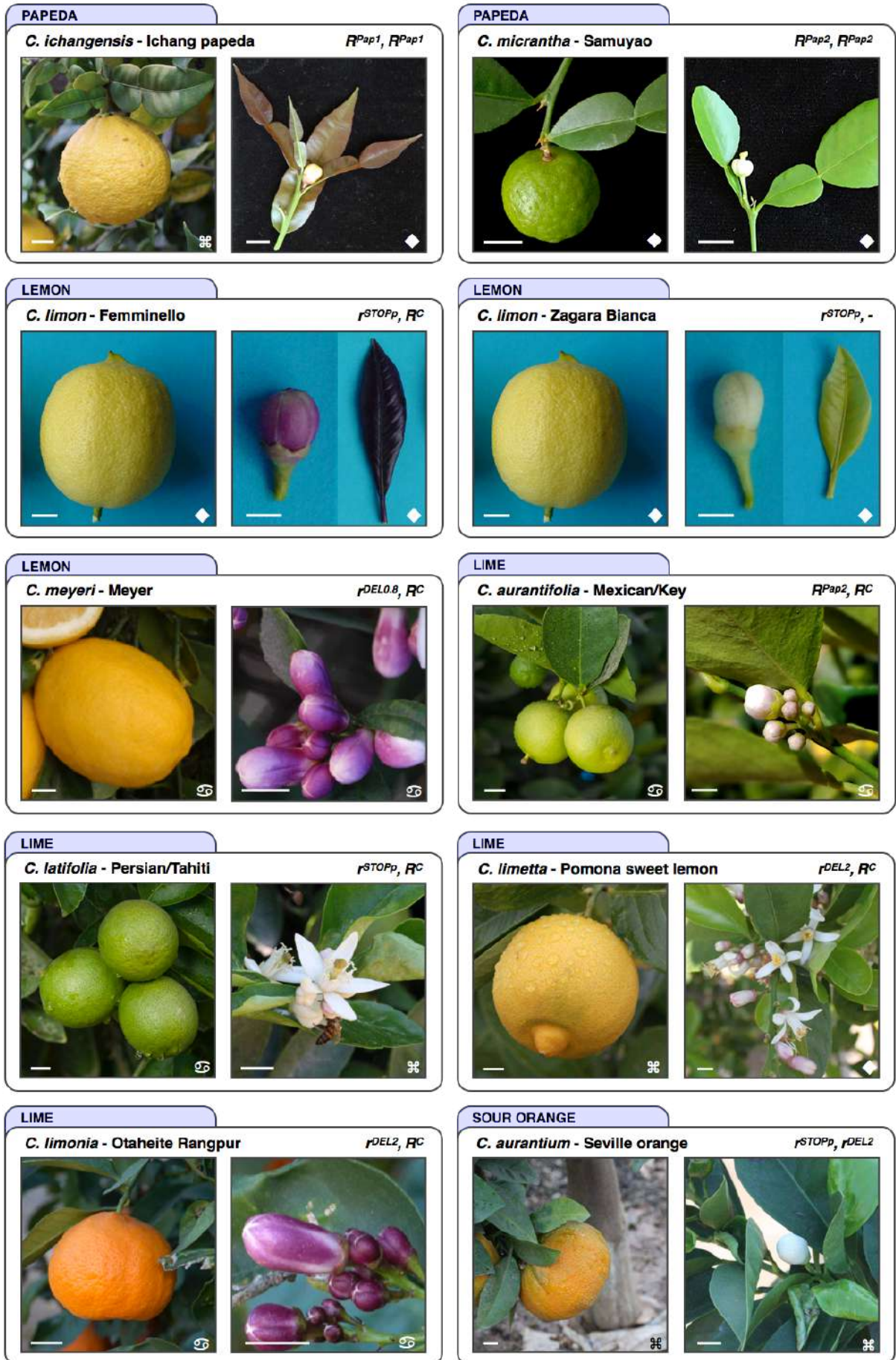
- 986 **Netzel M, Netzel G, Tian QG, Schwartz S, Konczak I** (2006) Sources of antioxidant  
987 activity in Australian native fruits. Identification and quantification of  
988 anthocyanins. *Journal of Agricultural and Food Chemistry* **54**: 9820-9826  
989
- 990 **Nicolosi E, Deng ZN, Gentile A, La Malfa S, Continella G, Tribulato E** (2000)  
991 Citrus phylogeny and genetic origin of important species as investigated by  
992 molecular markers. *Theoretical and Applied Genetics* **100**: 1155-1166  
993
- 994 **Penjor T, Anai T, Nagano Y, Matsumoto R, Yamamoto M** (2010) Phylogenetic  
995 relationships of Citrus and its relatives based on rbcL gene sequences. *Tree*  
996 *Genetics & Genomes* **6**: 931-939  
997
- 998 **Penjor T, Yamamoto M, Uehara M, Ide M, Matsumoto N, Matsumoto R, Nagano**  
999 **Y** (2013) Phylogenetic Relationships of Citrus and Its Relatives Based on matK  
1000 Gene Sequences. *Plos One* **8**: e62574  
1001
- 1002 **Quattrocchio F, Wing JF, Leppen HTC, Mol JNM, Koes RE** (1993) Regulatory  
1003 Genes-Controlling Anthocyanin Pigmentation Are Functionally Conserved  
1004 among Plant-Species and Have Distinct Sets of Target Genes. *Plant Cell* **5**:  
1005 1497-1512  
1006
- 1007 **Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R**  
1008 (1999) Molecular analysis of the anthocyanin2 gene of petunia and its role in  
1009 the evolution of flower color. *Plant Cell* **11**: 1433-1444  
1010
- 1011 **Ramadugu C, Pfeil BE, Keremane ML, Lee RF, Maureira-Butler IJ, Roose ML**  
1012 (2013) A Six Nuclear Gene Phylogeny of Citrus (Rutaceae) Taking into Account  
1013 Hybridization and Lineage Sorting. *Plos One* **8**: e68410  
1014
- 1015 **Ramsay NA, Glover BJ** (2005) MYB-bHLH-WD40 protein complex and the  
1016 evolution of cellular diversity. *Trends in Plant Science* **10**: 63-70  
1017
- 1018 **Sainsbury F, Thuenemann EC, Lomonossoff GP** (2009) pEAQ: versatile  
1019 expression vectors for easy and quick transient expression of heterologous  
1020 proteins in plants. *Plant Biotechnology Journal* **7**: 682-693  
1021
- 1022 **Sanchez R, Serra F, Tarraga J, Medina I, Carbonell J, Pulido L, de Maria A,**  
1023 **Capella-Gutierrez S, Huerta-Cepas J, Gabaldon T, Dopazo J, Dopazo H**  
1024 (2011) Phylemon 2.0: a suite of web-tools for molecular evolution,  
1025 phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Research*  
1026 **39**: W470-W474  
1027
- 1028 **SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL** (1998) The  
1029 paleontology of intergene retrotransposons of maize. *Nature Genetics* **20**: 43-  
1030 45  
1031
- 1032 **Schwinn K, Venail J, Shang YJ, Mackay S, Alm V, Butelli E, Oyama R, Bailey P,**  
1033 **Davies K, Martin C** (2006) A small family of MYB-regulatory genes controls  
1034 floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell*  
1035 **18**: 831-851

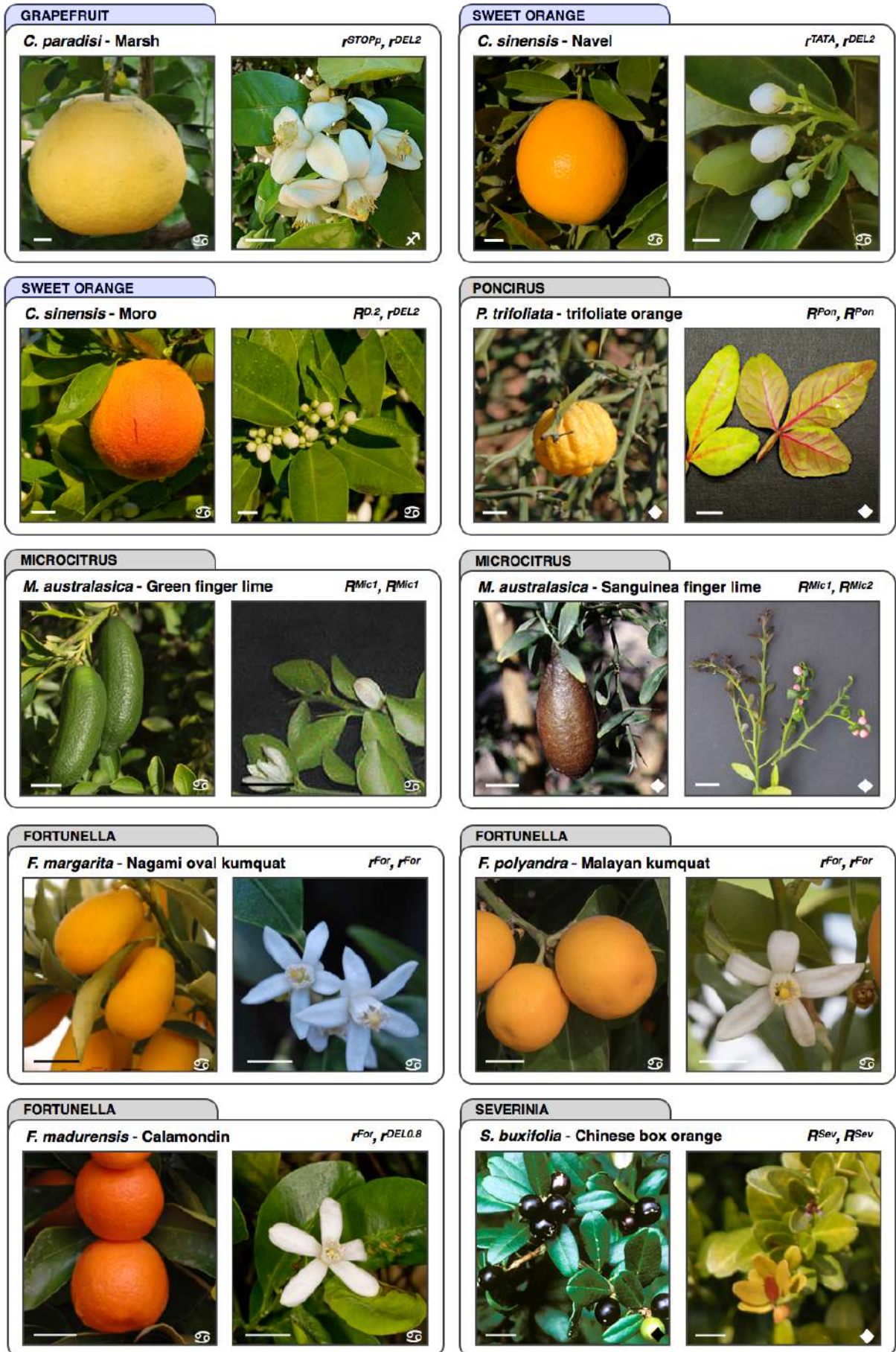
- 1036  
1037 **Scora RW** (1975) Symposium on Biochemical Systematics, Genetics and Origin of  
1038 Cultivated Plants .9. History and Origin of Citrus. Bulletin of the Torrey  
1039 Botanical Club **102**: 369-375  
1040  
1041 **Steyn WJ, Wand SJE, Holcroft DM, Jacobs G** (2002) Anthocyanins in vegetative  
1042 tissues: a proposed unified function in photoprotection. New Phytologist **155**:  
1043 349-361  
1044  
1045 **Stracke R, Werber M, Weisshaar B** (2001) The R2R3-MYB gene family in  
1046 Arabidopsis thaliana. Current Opinion in Plant Biology **4**: 447-456  
1047  
1048 **Swingle WT** (1946) The botany of Citrus and its wild relatives of the orange  
1049 subfamily (family Rutaceae, subfamily Aurantioideae). *In* W Reuther, H  
1050 Webber, L Batchelor, eds, The Citrus Industry. University of California Press,  
1051 Berkeley, CA, pp 129-474  
1052  
1053 **Swingle WT, Reece PC** (1967) The botany of Citrus and its wild relatives. *In* W  
1054 Reuther, H Webber, L Batchelor, eds, The Citrus Industry. University of  
1055 California Press, Berkeley, CA, pp 190–430  
1056  
1057 **Tamura K** (1992) The rate and pattern of nucleotide substitution in Drosophila  
1058 mitochondrial DNA. Mol Biol Evol **9**: 814-825  
1059  
1060 **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S** (2011) MEGA5:  
1061 Molecular Evolutionary Genetics Analysis Using Maximum Likelihood,  
1062 Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology  
1063 and Evolution **28**: 2731-2739  
1064  
1065 **Tanaka T** (1977) Fundamental discussion of citrus classification. Studia Citrologica  
1066 **14**: 1-6  
1067  
1068 **Tanaka Y, Sasaki N, Ohmiya A** (2008) Biosynthesis of plant pigments:  
1069 anthocyanins, betalains and carotenoids. Plant Journal **54**: 733-749  
1070  
1071 **Webber HJ** (1943) Plant characteristics and climatology. *In* W Reuther, H Webber, L  
1072 Batchelor, eds, The Citrus Industry, Vol 1. University of California Press,  
1073 Berkeley, CA, pp 41-69  
1074  
1075 **Winkel-Shirley B** (2001) Flavonoid biosynthesis. A colorful model for genetics,  
1076 biochemistry, cell biology, and biotechnology. Plant Physiology **126**: 485-493  
1077  
1078 **Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M,**  
1079 **Scalabrin S, Terol J et al.** (2014) Sequencing of diverse mandarin, pummelo  
1080 and orange genomes reveals complex history of admixture during citrus  
1081 domestication. Nature Biotechnology **32**: 656-662  
1082  
1083 **Xu Q, Chen LL, Ruan XA, Chen DJ, Zhu AD, Chen CL, Bertrand D, Jiao WB,**  
1084 **Hao BH, Lyon MP et al.** (2013) The draft genome of sweet orange (Citrus  
1085 sinensis). Nature Genetics **45**: 59-U92

1086  
1087  
1088  
1089  
1090  
1091

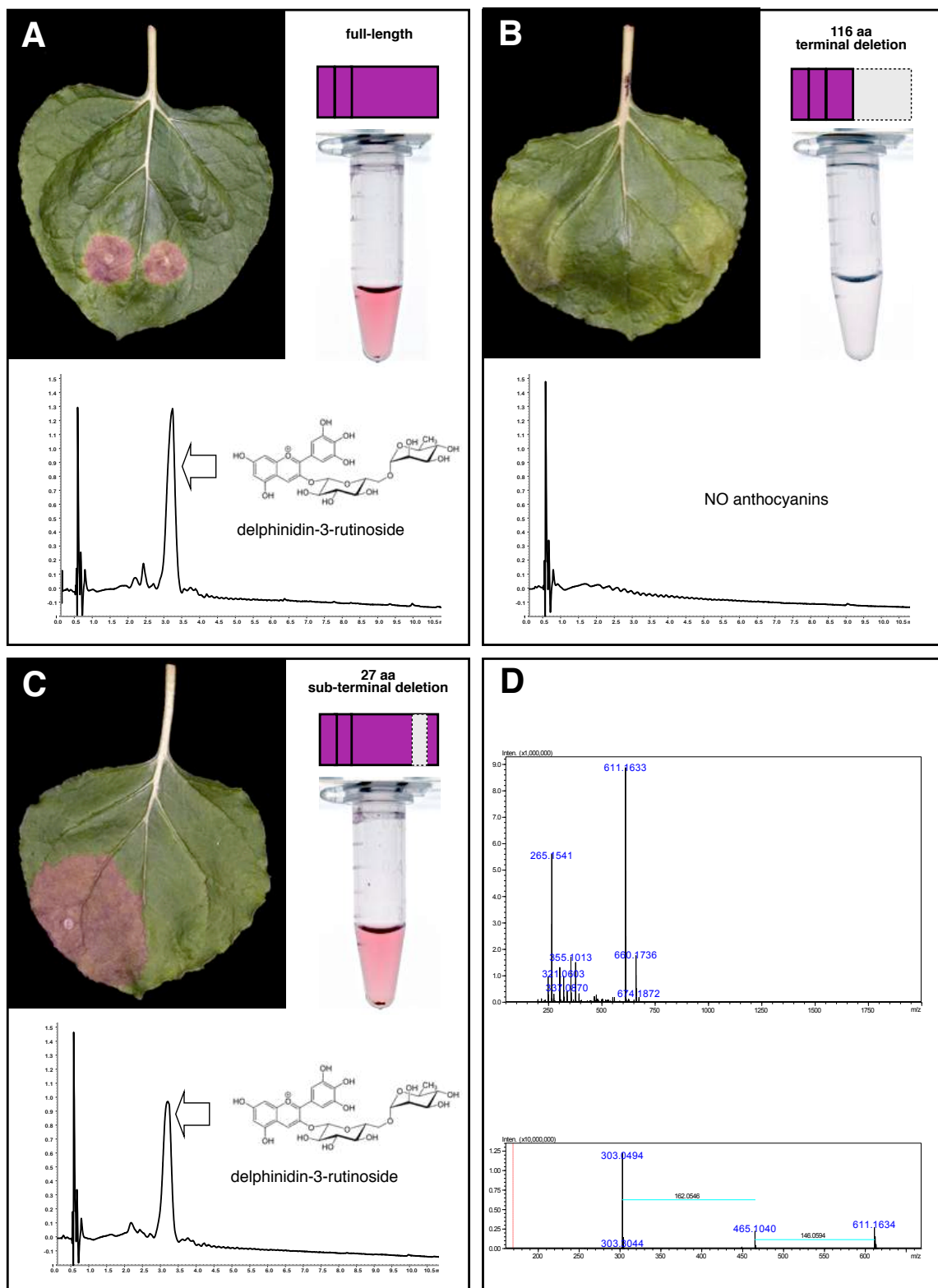






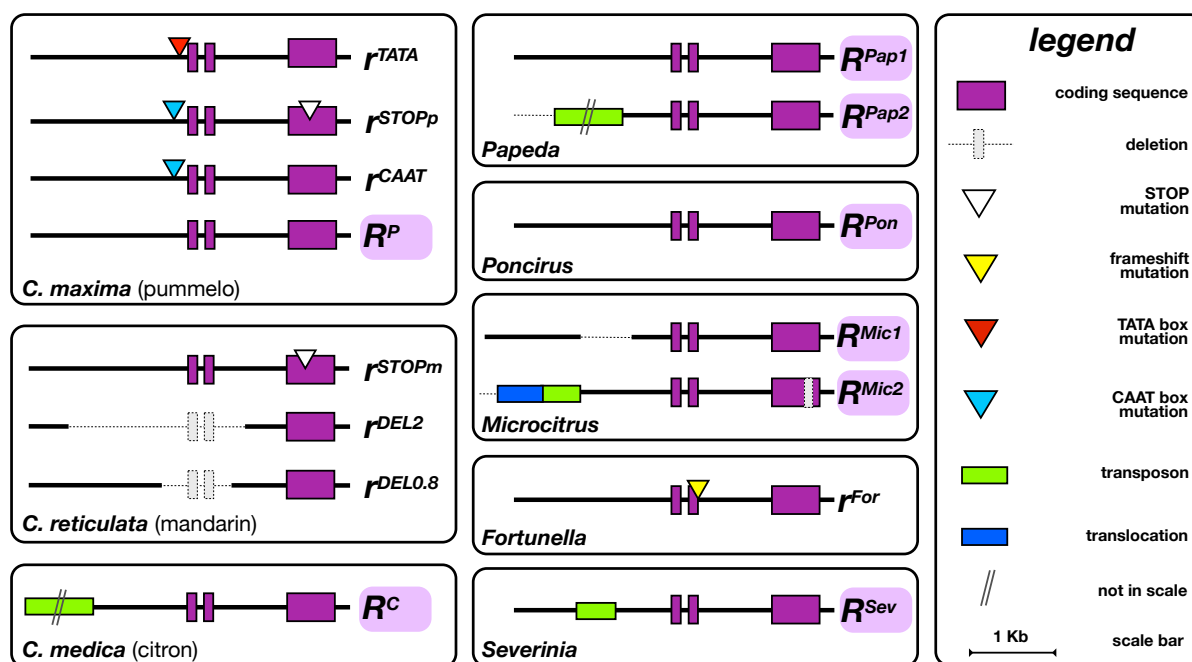


**Figure 1.** Phenotypes of some of the accessions considered in this study and listed in Table 1. For each accession, the common and the scientific name according to Tanka's classification are provided, together with the allelic constitution at the *Ruby* locus. Images of fruit and flowers or young leaves are presented to show the presence or absence of anthocyanin pigmentation. Scale bar is approximately 1 cm. The images were obtained from the following sources: ◆, pictures taken by the authors; ☞, UC-Riverside Citrus Variety Collection (<http://www.citrusvariety.ucr.edu/index.html>); ☚, Citrus ID Tools (<http://idtools.org/id/citrus/citrusid/>); ⚡, Radoslav Krchnavek; ⚡; Eve and George DeLange.

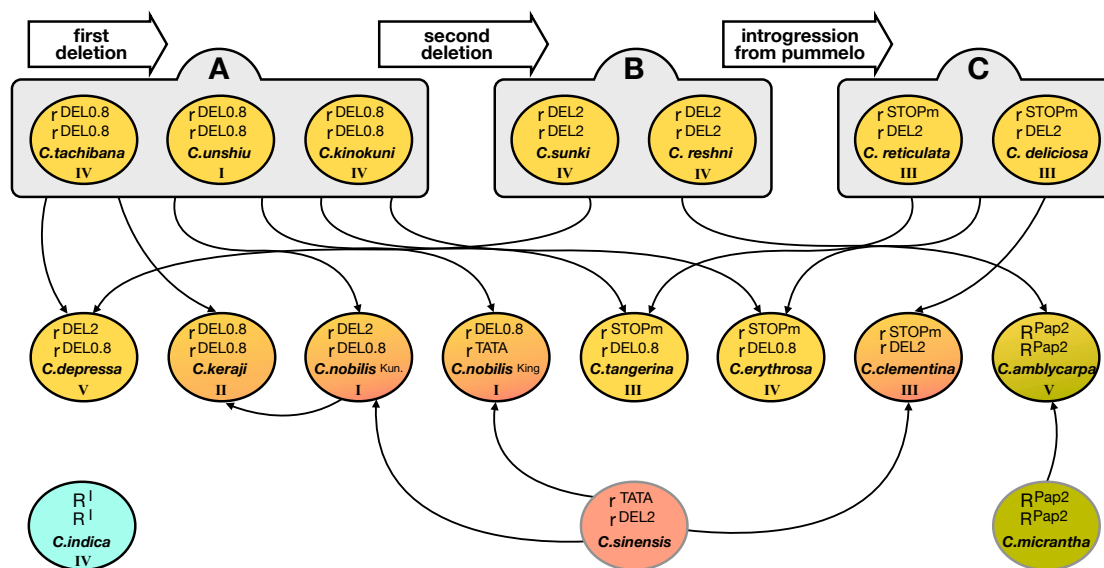


**Figure 2.**

**Figure 2.** Functionality of different *Ruby* alleles. *Nicotiana benthamiana* leaves were infiltrated with cDNA constructs corresponding to three different *Ruby* alleles. The phenotypes of leaves, the colour of the leaf extracts and the HPLC chromatograms at 525 nm 7 days 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 *Microcitrus australasica* var ‘Sanguinea’ encoding a protein with a 27 amino acid 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 4.25ppm). Note the fragment representing delphinidin (expected mass 303.0499) and the losses of a hexose moiety (expected loss 162.0528) and rhamnose moiety (expected loss 146.0579).

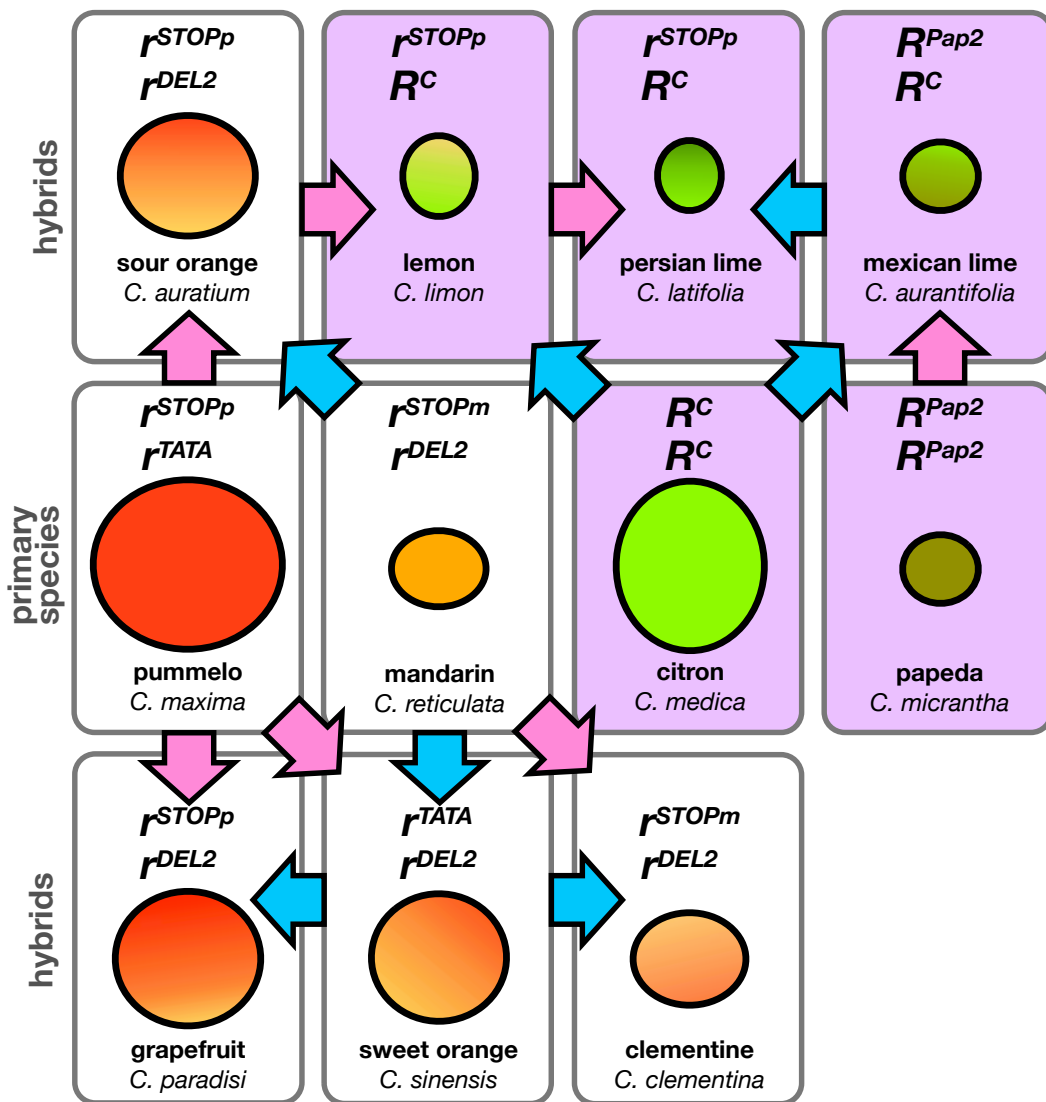


**Figure 3.** Schematic representation of different *Ruby* alleles identified in primary species of *Citrus* and related genera. Functional alleles (*R*) associated with anthocyanin pigmentation are highlighted in purple.

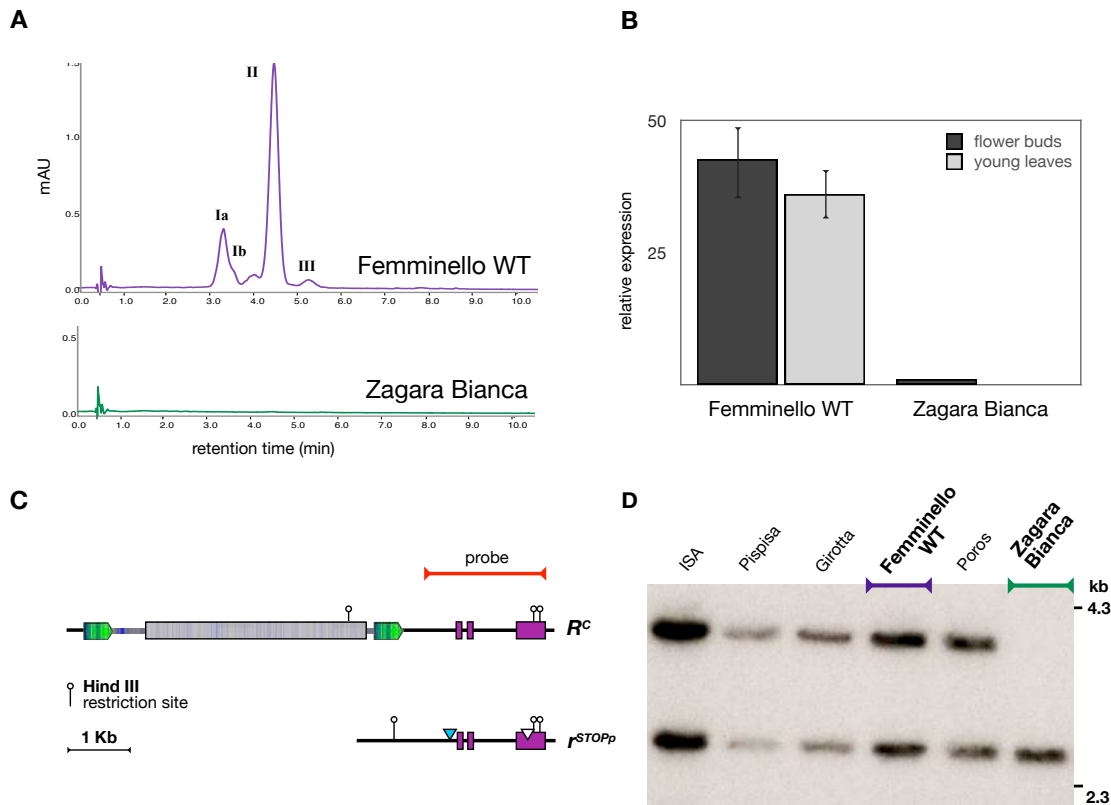


**Figure 4.** Classification of mandarins based on the analysis of the *Ruby* gene. Mandarins can be classified into three groups (A, B and C) on the basis of two deletions and an introgression from pummelo. Many mandarin varieties are hybrids between members of these three groups or interspecific hybrids involving other *Citrus* species. *Citrus 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 colour shading indicates the contribution from genomes of other species. Common names for the mandarin accessions are provided in Table 1.

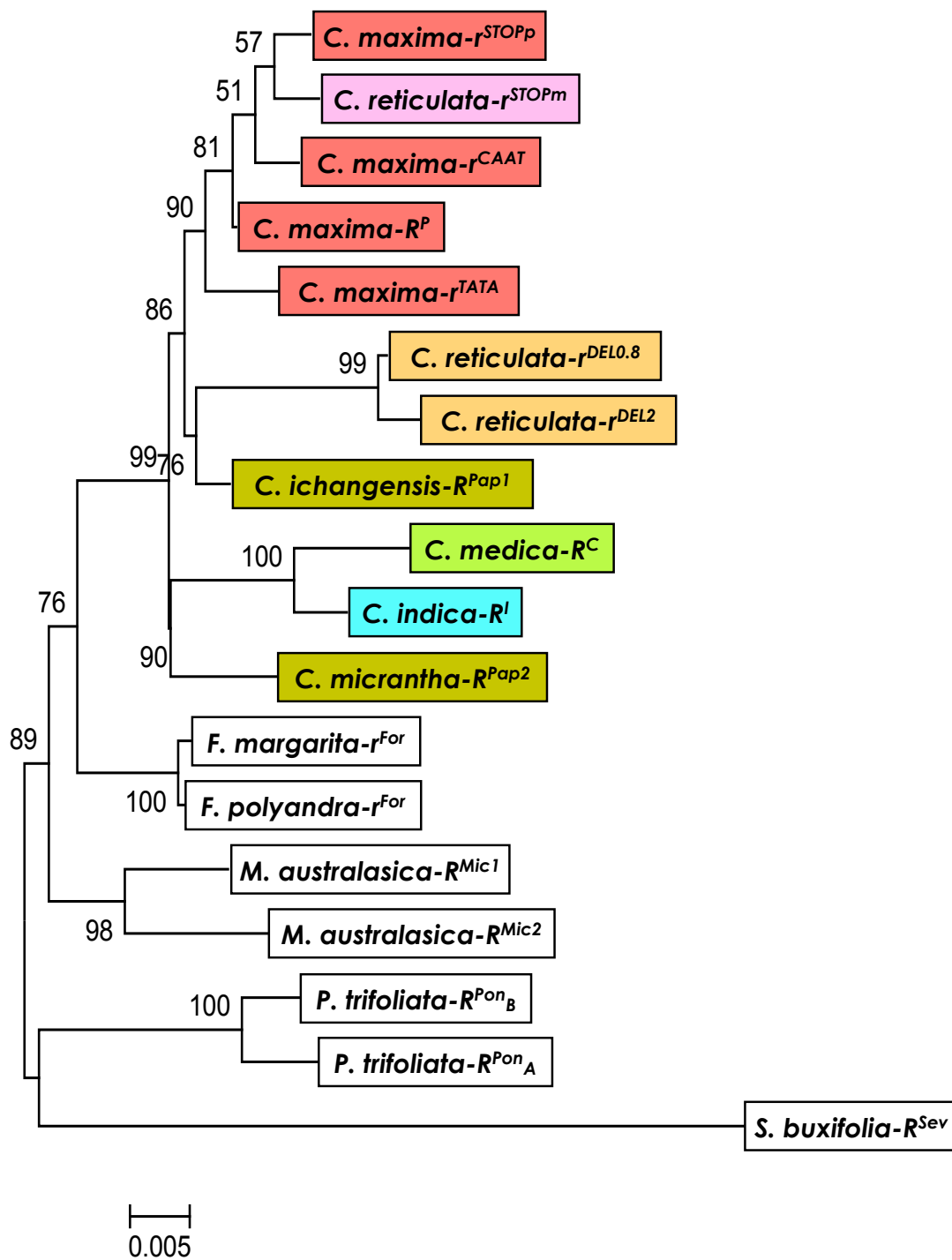




**Figure 5.** Segregation analysis of *Ruby* in some of the commercially important *Citrus* species and hybrids considered in this study. Simplified diagram showing a single existing variety of pummelo and mandarin as potential parents of both sweet and sour orange, which is highly unlikely. Light purple 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 or parental contribution to the generation of a hybrid based on specific mitotypes and chlorotypes as determined in previously studies. The gradation in colour shading indicates the contribution from different species to the derivation of the hybrids.



**Figure 6.** Analysis of wild type and *white* mutant of lemon. **A**, HPLC chromatogram 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)- $\beta$ -glucoside; **III**, peonidin 3-(6"-malonylglucoside). Identification of peaks is based on UHPLC-ESI-MS analysis (Supplemental Fig. S8, S9 and S10) and comparison to literature data. **B**, Relative qRT-PCR quantification of *Ruby* expression in flower buds and young leaves of ‘Femminello’ and ‘Zagara Bianca’ lemons. Error bars represent SE of the mean. **C**, Schematic diagram of the two alleles of *Ruby* identified in lemon,  $R^C$  and  $r^{STOPp}$ . The different symbols are explained in Fig. 3 and Supplemental Fig. S2A. The *Hind*III 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 *Hind*III and probed with a  $^{32}$ P-labelled probe of the *Ruby* gene. ‘Zagara Bianca’, the only variety completely unable to produce anthocyanins, doesn’t display the hybridization band corresponding to the active allele  $R^C$ .



**Figure 7.** Phylogenetic relationship between primary species of *Citrus* and related genera based on the sequence of *Ruby*. The phylogenetic trees were generated using Phylemon software program. Branch support is given in all branches. The different *Ruby* alleles are colour-coded as in Table 1. The sequences of the *Ruby* gene and its promoter used for phylogenetic analysis have been deposited in GenBank under accession numbers KT591672-KT591689.

**Table 1. Species and hybrids used in this study**

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

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

*Citrus* accessions were obtained from the following sources: **A**: IVIA, Instituto Valenciano de Investigaciones Agrarias, Moncada, Valencia, Spain; **B**: CREA-CRA, Centro di Ricerca per l'Agricoltura e le Colture Mediterranee, Acireale, Italy; **C**: USDA-ARS National Clonal Germplasm Repository for Citrus & Dates, Riverside, CA, USA; **D**: CRB CITRUS, INRA-CIRAD, Citrus Biological Resource Center, San Giuliano, Corsica, France; **E**: Huazhong Agricultural University, Wuhan, China; **F**: Jingzhou, Huannan Province, China; **G**: Waitrose Supermarket, Norwich, UK; **H**: Notcutts Garden Centre, Norwich, UK.

In accessions of the genus *Citrus*, the *Ruby* alleles are colour-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 coloured. The lemon accession indicated with an asterisk (\*) is a 'graft chimera' and contains a third *Ruby* allele that may have originated from an acidless limetta. Pictures of fruit and flowers of the accessions indicated in bold are provided in Fig.1.

## Parsed Citations

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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. American Journal of Botany 96: 668-685**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Carbonell-Caballero J, Alonso R, Ibanez 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. Molecular Biology and Evolution 32: 2015-2035**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Coen ES, Carpenter R, Martin C (1986) Transposable Elements Generate Novel Spatial Patterns of Gene-Expression in Antirrhinum Majus. Cell 47: 285-296**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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 Genetics 15: 152**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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. Annals of Botany 117: 565-583**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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 Journal 49: 414-427**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Fabroni S, Ballistreri G, Amenta M, Rapisarda P (2016) Anthocyanins in different Citrus species: an UHPLC-PDA-ESI/MSn-assisted qualitative and quantitative investigation. Journal of the Science of Food and Agriculture 96: 4797-4808**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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. Theoretical and Applied Genetics 96: 812-822**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Fincham JRS (1967) Mutable Genes in Light of Callans Hypothesis of Serially Repeated Gene Copies. Nature 215: 864-866**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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 Genetics & Genomes 7: 49-61**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Garcia-Lor A, Ancillo G, Navarro L, Ollitrault P (2013) Citrus (Rutaceae) SNP Markers Based on Competitive Allele-Specific PCR; Transferability across the Aurantioideae Subfamily. Applications in Plant Sciences 1: 1200406**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Garcia-Lor A, Curk F, Snoussi-Trifa H, Morillon R, Ancillo G, Luro F, Navarro L, Ollitrault P (2013) 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. Annals of Botany 111: 1-19**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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 Genetics & Genomes 11: 123**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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, CA, pp 431-591**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Kimura M (1980) A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide-Sequences. Journal of Molecular Evolution 16: 111-120**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Kitajima A, Yamasaki A, Habu T, Preedasuttijit B, Hasegawa K (2007) Chromosome identification and karyotyping of satsuma mandarin by genomic in situ hybridization. Journal of the American Society for Horticultural Science 132: 836-841**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

Downloaded from on October 5, 2020 - Published by www.plantphysiol.org  
Copyright © 2017 American Society of Plant Biologists. All rights reserved.

CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Krueger R, Navarro L (2007) Citrus germplasm resources. In IA Khan, ed, Citrus genetics, breeding, and biotechnology. CAB International, Wallingford, UK, pp 45-140**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Lai YT, Chen IZ (2008) Effect of Temperature on Calamondin (Citrus microcarpa) Flowering and Flower Bud Formation. Proceedings of the International Symposium on Citrus and Other Tropical and Subtropical Fruit Crops: 111-115**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Liu CY, Wang X, Xu YT, Deng XX, Xu Q (2014) Genome-wide analysis of the R2R3-MYB transcription factor gene family in sweet orange (Citrus sinensis). Molecular Biology Reports 41: 6769-6785**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu JY, Stamatakis AP, Linder CR (2012) SATE-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees. Systematic Biology 61: 90-106**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Malinowski E (1935) Studies on Unstable characters in Petunia. The extreme flower types of the Unstable race with mosaic color patterns. Genetics 20: 342-356**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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 Biology 8: 83**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**McClintock B (1984) The Significance of Responses of the Genome to Challenge. Science 226: 792-801**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Moore T (1866) The Treasury of Botany: a popular dictionary of the vegetable kingdom, Vol 1. Longmans Green & Co., London**

**Morton JF (1987) Grapefruit. In Fruits of warm climates. University of Florida, Miami, pp 152-158**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Netzel M, Netzel G, Tian QG, Schwartz S, Konczak I (2006) Sources of antioxidant activity in Australian native fruits. Identification and quantification of anthocyanins. Journal of Agricultural and Food Chemistry 54: 9820-9826**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Nicolosi E, Deng ZN, Gentile A, La Malfa S, Contarella G, Tribulato E (2000) Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theoretical and Applied Genetics 100: 1155-1166**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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



Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Quattrocchio F, Wing JF, Leppen HTC, Mol JNM, 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**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Sainsbury F, Thuenemann EC, Lomonosoff GP (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnology Journal 7: 682-693**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Sanchez R, Serra F, Tarraga J, Medina I, Carbonell J, Pulido L, de Maria A, Capella-Gutierrez S, Huerta-Cepas J, Gabaldon T, Dopazo J, Dopazo H (2011) Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Research 39: W470-W474**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Schwinn K, Venail J, Shang YJ, 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**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Scora RW (1975) Symposium on Biochemical Systematics, Genetics and Origin of Cultivated Plants .9. History and Origin of Citrus. Bulletin of the Torrey Botanical Club 102: 369-375**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. Current Opinion in Plant Biology 4: 447-456**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)  
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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, CA, pp 129-474**

Pubmed: [Author and Title](#)  
CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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, CA, pp 190-430**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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, CA, pp 41-69**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**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. Nature Biotechnology 32: 656-662**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

**Xu Q, Chen LL, Ruan XA, Chen DJ, Zhu AD, Chen CL, Bertrand D, Jiao WB, Hao BH, Lyon MP et al. (2013) The draft genome of sweet orange (*Citrus sinensis*). Nature Genetics 45: 59-U92**

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)