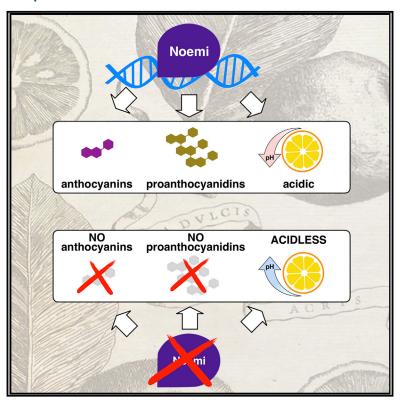
Current Biology

Noemi Controls Production of Flavonoid Pigments and Fruit Acidity and Illustrates the Domestication Routes of Modern Citrus Varieties

Graphical Abstract



Authors

Eugenio Butelli, Concetta Licciardello, Chandrika Ramadugu, ..., Giuseppe Reforgiato Recupero, Yann Froelicher, Cathie Martin

Correspondence

eugenio.butelli@jic.ac.uk

In Brief

In some varieties of citrus, exceptionally low fruit acidity is associated with absence of anthocyanin pigments in leaves and flowers and proanthocyanidins in seeds. Taking advantage of natural variation, Butelli et al. show that this pleiotropic phenotype is the result of mutations in a single gene, *Noemi*, encoding a bHLH transcription factor.

Highlights

- Noemi is essential for the production of flavonoid pigments in citrus
- Noemi is essential for the regulation of fruit acidity in citrus
- Retrotransposons are associated with the acidless phenotype in commercial varieties
- A specific ancient mutation retraces the steps of citrus history and cultivation





Noemi Controls Production of Flavonoid Pigments and Fruit Acidity and Illustrates the Domestication Routes of Modern Citrus Varieties

Eugenio Butelli, ^{1,7,*} Concetta Licciardello, ² Chandrika Ramadugu, ³ Marie Durand-Hulak, ⁴ Alessandra Celant, ⁵ Giuseppe Reforgiato Recupero, ² Yann Froelicher, ⁶ and Cathie Martin ¹

¹John Innes Centre, Norwich NR4 7UH, UK

²CREA-OFA, Research Centre for Olive, Citrus, and Tree Fruit, Corso Savoia 190, 95024 Acireale, Italy

³University of California, Riverside, Riverside, CA 92521, USA

⁴INRA, Unité Mixte de Recherche AGAP, Station Institut National de la Recherche Agronomique, 20230 San Giuliano, France

⁵Laboratory of Palaeobotany and Palynology, Department of Environmental Biology, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy

⁶CIRAD Unité Mixte de Recherche AGAP, Station Institut National de la Recherche Agronomique, 20230 San Giuliano, France ⁷Lead Contact

*Correspondence: eugenio.butelli@jic.ac.uk https://doi.org/10.1016/j.cub.2018.11.040

SUMMARY

In citrus, the production of anthocyanin pigments reguires the activity of the transcriptional activator Ruby. Consequently, loss-of-function mutations in Ruby result in an anthocyaninless phenotype [1]. Several citrus accessions, however, have lost the ability to produce these pigments despite the presence of wild-type Ruby alleles. These specific mutants have captivated the interest of botanists and breeders for centuries because the lack of anthocyanins in young leaves and flowers is also associated with a lack of proanthocyanidins in seeds and, most notably, with an extreme reduction in fruit acidity (involving about a three-unit change in pH). These mutants have been defined collectively as "acidless" [2-4]. We have identified *Noemi*, which encodes a basic helix-loop-helix (bHLH) transcription factor and which controls these apparently unrelated processes. In accessions of Citron, limetta, sweet lime, lemon, and sweet orange, acidless phenotypes are associated with large deletions or insertions of retrotransposons in the *Noemi* gene. In two accessions of limetta, a change in the core promoter region of Noemi is associated with reduced expression and increased pH of juice, indicating that Noemi is a major determinant of fruit acidity. The characterization of the Noemi locus in a number of varieties of Citron indicates that one specific mutation is ancient. The presence of this allele in Chinese fingered Citrons and in those used in the Sukkot Jewish ritual [5] illuminates the path of domestication of Citron, the first citrus species to be cultivated in the Mediterranean. This allele has been inherited in Citron-derived hybrids with long histories of cultivation.

RESULTS AND DISCUSSION

Citrus is a complex group of flowering plants with notable nutritional, medicinal, and aromatic value. Citrus became one of the world's most economically important fruit crops through a largely obscure history of evolution and domestication. The genus originated in a wide area spanning North-Eastern India to South China and South-East Asia and is thought to have been cultivated for at least 3,300 years [6–8].

Within citrus, the existence of a "syngameon"-a group of genetically related organisms linked by interspecific hybridization [9]—is more appropriate than distinguishing individual species, since marked sexual compatibility between species has generated hundreds of citrus varieties, often with broad morphological diversity [10]. However, genetic studies have identified Pummelo (C. maxima), Citron (C. medica), and Mandarin (C. reticulata) as the fundamental or primary species [11, 12] (Figure 1), with a fourth species from Papeda (C. micrantha) involved in the origin of some limes [13]. Other commercially important citrus are the result of hybridization among these four primary species [14]. We have used anthocyanin production to facilitate our understanding of the taxonomy and domestication of this important group of fruit crops. Among primary species, Citron produces anthocyanins naturally in its leaves and young flower buds, whereas most Pummelo and all Mandarin accessions do not, due to mutations affecting the expression or activity of the Ruby gene, which encodes an R2R3MYB transcription factor [1, 15]. The ability of hybrids to make anthocyanins depends on the functionality of the Ruby alleles inherited from their parental species [1]. The hybrids are able to maintain their unique genetic composition because they are propagated clonally, either by apomictic seeds or by grafting.

While almost all anthocyanin production in citrus can be explained by variations in the activity of the *Ruby* gene, one specific subset of mutants has lost anthocyanin pigmentation despite the presence of wild-type *Ruby* alleles. This group of anthocyaninless variants is characterized by two accompanying traits: the



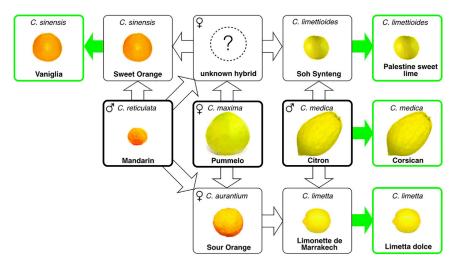


Figure 1. Genetic Origin of the Citrus Accessions Considered in This Study

Maternal and paternal contribution of the three primary species Citron, Pummelo, and Mandarin (boxed in bold) in the generation of acidless varieties of citrus. Green arrows indicate mutations resulting in the acidless phenotype. The precise origin of sweet orange remains unknown.

absence of proanthocyanidins in the seeds (apparent phenotypically as the absence of the "chalazal spot," a characteristically colored round area on the inner seed coat) and fruit with juice almost completely devoid of acidity [2–4]. The invariant combination of these three traits (low fruit acidity, white flowers and green leaves, seeds of light cream color) defines the so-called "acidless" phenotype (Figure 2).

Acidless varieties are often defined as "sweet," reflecting their insipid flavor and the sharp increase in the sugar-acid ratio in their fruit, which is the most important determinant of fruit quality and taste in citrus. The multifaceted anthocyaninless, proanthocyanidin-less, acidless phenotype is likely to be the result of mutations in a single gene. From 33 citrus varieties completely unable to produce anthocyanins, acidless varieties are the only accessions containing functional alleles of *Ruby*, a key regulatory *MYB* gene essential for anthocyanin production [1], suggesting strongly for pleiotropic effects of mutations in a single gene.

In angiosperms, anthocyanin biosynthesis is regulated by conserved MYB/bHLH/WD40 (MBW) complexes, formed by the interaction between MYB transcription factors, basic helixloop-helix (bHLH) transcription factors, and WD40-repeat proteins [16, 17]. The three elements of the complex act in pyramidal fashion. While the MYB factor provides the DNA-binding specificity for the activation of the target genes, the other two components are often involved in regulating additional processes including proanthocyanidin biosynthesis, which involves an MBW complex with the same bHLH transcription factor that regulates anthocyanin biosynthesis and a specific R2R3MYB protein that determines specificity for target genes. Pivotal studies in Petunia have linked regulation of anthocyanin biosynthesis and vacuolar acidification through a common bHLH protein (AN1) working in the complex [18-20]. Consequently, we tested whether members of the MBW complex might be responsible for the acidless phenotypes in citrus by searching the genome of C. clementina (https://phytozome.jgi.doe.gov) for potential citrus homologs of key regulatory genes in Petunia and Arabidopsis.

We first considered an acidless Citron mutant because Citron is a true species with a low level of heterozygosity that can facilitate genomic analysis. A wild-type Citron, "Poncire commun," can synthesize both anthocyanins and proanthocyanidins and

has a normal juice acidity of around pH 2.5. In contrast, "Corsican" Citron is completely unable to produce flavonoid pigments, either anthocyanins in its leaves and flowers or proanthocyanidins in its seeds, and has very low juice acidity around pH 5.5 (Table 1). Poncire commun is the ideal control for Corsican Citron, since it is very closely genetically related

[21] (Figure 2). For both accessions, we sequenced *Ruby* (corresponding to Ciclev10013455 in the *C. clementina* genome), two candidate genes encoding proteins with homology to the WD-40 proteins of the MBW complex (Ciclev10005375 and Ciclev10015815), the *bHLH* gene *MYC2* (Ciclev10019219) [22], and an uncharacterized gene (Ciclev10019118) encoding a bHLH protein closely related to TT8 in *Arabidopsis* and AN1 in Petunia, which we named *Noemi*.

While the sequences of the WD-40, genes *Ruby* and *MYC-2*, were identical in both the Corsican and Poncire commun accessions of Citron, PCR analysis indicated the presence of a deletion of 1,313 nucleotides in the 3' region of *Noemi* (Figures 3A and S1A). The deletion was homozygous in Corsican Citron $(n^{DEL3'}; n^{DEL3'})$ and heterozygous in Poncire commun $(N^C; n^{DEL3'})$. The wild-type allele (N^C) contains seven introns and encodes a protein of 695 amino acids. The mutation in $n^{DEL3'}$ results in deletion of the sequences encoding the last 275 amino acids and is predicted to generate a non-functional truncated protein lacking the entire bHLH domain required for dimerization and interaction with other proteins [23, 24] (Figure S1B).

Unlike most cultivated citrus, Citron produces zygotic monoembryonic seeds, offering the opportunity for segregation analysis using seeds from the heterozygous Poncire commun. PCR analysis indicated that five seedlings, showing unequivocal production of anthocyanins, contained at least one wild-type *Noemi* allele, while the green seedlings were homozygous for the deletion ($n^{DEL3'}$) (Figure S1A).

Flowers or fruit have not yet been obtained from these seedlings because of the long juvenile period. However, a few plants were grown from seeds of self-pollinated Poncire commun in 2013 in Corsica. Two plants were kept: one with purple leaves with pigmented flowers that produced fully acidic fruit and seeds with proanthocyanidins; the other produced no anthocyanins nor seed proanthocyanidins and bore acidless fruit. PCR analysis confirmed that the *Noemi* deletion allele, $n^{DEL3'}$, was homozygous in the plant with the acidless phenotype.

Citron was the first citrus species to spread westward and reach the Mediterranean basin in the early Roman period [6]. Acidless fruit were certainly known during the Renaissance period [2] (Figure S2). Recent studies have suggested that





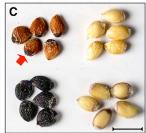


Figure 2. Wild-Type and Acidless Citrons (C. medica)

(A) Young leaves and flowers of Poncire commun (left) and Corsican Citron (right), a variety unable to produce anthocyanins.

(B) Germinating seeds in a fruit of Corsican Citron. The seeds are clear colored, the seedlings are anthocyaninless, and the juice is acidless.

(C) Seeds of Poncire commun (left) and Corsican Citron (right) before (upper panel) and after (lower panel) staining with p-Dimethylaminocinnamaldehyde (DMACA) reagent, indicating presence or absence of proanthocyanidins. The red arrow indicates the chalazal spot (scale bars, 1 cm).

See also Figures S1 and S2 and Table S1.

Chinese Citron varieties are genetically distinct from those that underwent a wave of diversification in the Mediterranean region [25]. The $n^{DEL3'}$ allele might represent a recent mutation, consistent with the view of Hodgson that Corsican Citron originated in Corsica [4]. However, we found that another acidless variety, "Assads Moroccan" Citron [3], is also homozygous for the $n^{DEL3'}$ allele, suggesting a common origin with Corsican Citron despite phenotypic and genetic differences [25]. More remarkably, the "Yemen" Citron, a very ancient pulpless variety without juice sacs used in the Sukkot Jewish tradition since the time of the destruction of the First Temple, is also heterozygous for the $n^{DEL3'}$ allele. Another variety traditionally used for this religious ritual, "Greek" Citron, has the same heterozygous constitution at the Noemi locus (Figure S1C; Table S1). Historically, these findings support claims of the importance of religion and Jewish communities on the spread of Citron in the Mediterranean region [26, 27]. It also suggests that the "authentic" Jewish Citron, or Etrog (nongrafted and not hybridized with other varieties) could have been an acidless one, an idea supported by a reference to "sweet Citron" in the Talmud (200 AD) [28]. This probably refers to an acidless Citron, because alternative sweeter citrus varieties did not arrive in the Middle East until the 10th century and were introduced to Europe at the end of the 15th century [6].

We examined the Noemi locus in a number of fingered varieties of Citron (C. medica var. sarcodactylis), characterized by fruit segmented into finger-like sections (Figure S1D). While usually considered as a single variety under the name of "Buddha's Hand," many distinct varieties are known in Asia [5, 25]. Fingered Citrons have no juicy pulp or seeds and represent "genotypes frozen in time." Both the common Buddha's Hand and a Japanese fingered accession were heterozygous (N^{C} ; $n^{DEL3'}$) at the Noemi locus. More interestingly, two related Chinese fingered Citron varieties, "Qingpi" and "Aihua," used in traditional medicine, characterized by white flowers and green leaves [5], were homozygous for the $n^{DEL3'}$ allele (Figure S1C; Table S1). This confirmed that Noemi is essential for anthocyanin production and that the $n^{DEL3'}$ allele originated before the arrival of Citron in the Mediterranean, suggesting that it may have been inherited by some of the many hybrids derived from Citron [13].

One such hybrid is "Palestine sweet lime," a true acidless variety still popular in India where it probably originated. The corresponding acidic form, which produces flavonoid pigments, is the rare "Soh Synteng" [4]. Palestine sweet lime and Soh Synteng (both also referred to as C. limettioides) are likely hybrids between Citron and a Pummelo \times Mandarin hybrid [13] (Figure 1). We determined the allelic composition of *Noemi* in five acidless accessions and in Soh Synteng. All the accessions contained the same mutated $n^{DEL3'}$ allele, derived from Citron. Soh Synteng contained a second *Noemi* allele, which appears, from its sequence, to be functional while all the acidless accessions carried a second mutated allele of *Noemi*, $n^{DEL5'}$, involving a deletion of 741 bp spanning the 5' UTR, the first exon, and most of the second exon of the *Noemi* gene (Figures 3A and S3A).

Another interesting hybrid of Citron is *C. limetta*, a sour orange \times Citron cross [13] (Figure 1). The acidless form, unable to produce flavonoid pigments, has a long history of cultivation as depicted in Italian books and paintings of the XVII and XVIII centuries (Figures 3C and S2D). The corresponding acidic variety, "Limonette de Marrakech," is less well known [29]. Both varieties contain the non-functional allele of *Noemi* derived from Citron, $n^{DEL3'}$ (Figure S3B), but the functional N^P allele found in Limonette de Marrakech has been disrupted by the insertion of a retrotransposon, Tcl5, in the acidless "Limetta dolce" (n^{Tcl5} ; $n^{DEL3'}$; Figure 3A). Tcl5 is a recently inserted and potentially active retroelement, as indicated by its identical long terminal repeats (LTRs) and intact coding region.

Lemon (*C. limon*) is the most widely distributed Citron-derived hybrid (arising from a sour orange × Citron cross). In regular acidic lemons, both *Noemi* alleles are wild-type and presumably functional (*N*^M; *N*^C); Sweet varieties of lemon, unable to produce anthocyanins, have been described. All the varieties we have analyzed (often displaying variegated phenotypes, in terms of color and acidity, on the same plant; Figure S3C) contained new alleles not present in standard lemons (Figures S3B and S3D) and thought to be from periclinal graft chimeras [30, 31] where the L1 layer, which produces the epidermis and epidermal juice sacs of the fruit, came from an acidless limetta, while L2 and L3 came from lemon. Sweet lemons are not true lemons but chimeric accessions as observed in the XIX century [32] and reported in more recent studies [1, 32, 33].

Common Name	Tanaka's Classification	Juice pH	Anthocyanins	Proanthocyanidins	Source	Accession
Citron						
Poncire commun	C. medica	2.45	yes	yes	D	SRA701
Corsican	C. medica	5.42	no	no	D	SRA613
Diamante	C. medica	2.46	yes	yes	В	Palazzelli Certif.
Buddah's hand	C. medica	NA	yes	NA	С	CRC3768
Sweet lime						
Soh Synteng	C. limettioides	2.58	yes	yes	С	CRC3261
Palestine sweet lime	C. limettioides	5.41	no	no	С	CRC1482
Mary Ellen	C. limettioides	5.51	no	no	С	CRC4053
Unnamed	C. limettioides	5.39	no	no	С	CRC921
Unnamed	C. limettioides	5.54	no	no	С	CRC919
Unnamed	C. limettioides	5.32	no	no	С	CRC363
Limetta						
Limonette de Marrakech	C. limetta	2.55	yes	yes	C,D	CRC3989; SRA829
Limetta dolce	C. limetta	5.88	no	no	В	C-F2P7
Pomona	C. limetta	4.59	yes	yes	С	CRC4068
Millsweet	C. limetta	4.5	yes	yes	С	CRC569
Sweet Orange		'	'			
Navel	C. sinensis	3.47	no	yes	В	2B-F1-P14
Valencia	C. sinensis	3.34	no	yes	В	6A2-F5P1
Tarocco comune	C. sinensis	3.61	yes	yes	В	4-F20P9
Moro	C. sinensis	3.50	yes	yes	В	8-F1P2
Tarocco Ferreri	C. sinensis	6.32	yes	yes	В	2B-F14P14
Vaniglia biondo	C. sinensis	6.02	no	no	В	2A-F19P6
Vaniglia sanguigno	C. sinensis	6.74	no	no	В	1A-F7P10
Lemon						
JIC lemon	C. limon	2.42	yes	yes	А	S72
Politi apireno	C. limon	2.39	yes	NA	В	C-F3P1
Girotta (red leaf)	C. limon	3.20	yes	yes	В	C-F22P9
Girotta (green leaf)	C. limon	5.70	no	no	В	C-F22P9
Poros (red leaf)	C. limon	2.78	yes	yes	В	C-F41P7
Poros (green leaf)	C. limon	5.69	no	no	В	C-F41P7
Pispisa (red leaf)	C. limon	2.41	yes	yes	В	C-F41P1
Pispisa (green leaf)	C. limon	5.68	no	no	В	C-F41P1
ISA (red leaf)	C. limon	2.30	yes	yes	В	C-F29P5
ISA (green leaf)	C. limon	5.36	no	no	В	C-F29P5

The pH of fruit juices and the presence or absence of flavonoid pigments are indicated. Citrus accessions were obtained from the following sources: A: John Innes Centre, Norwich, UK; B: CREA-OFA, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia agraria, Olivicoltura Frutticoltura Agrumicoltura, 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. NA, not available because of absence of seeds or juice in the accession. See also Figures 3 and S1–S4 and Table S1.

Sweet orange (*C. sinensis*) is a complex Mandarin × Pummelo hybrid with no contribution from Citron. It contains two potentially functional *Noemi* alleles, both derived from the Mandarin genetic pool [34], which are expressed in wild-type fruit (Figure S4A). Sweet orange does not produce anthocyanins because of different mutations in both alleles of *Ruby* [1]. Despite the absence of anthocyanins, true acidless varieties can be identified based on seed color [4, 35]. Specific staining indicated that the light color of seeds is due to the absence of proanthocyanidins (Figure S4B). Two related acidless varieties, "Vaniglia bio-

ndo" and "Vaniglia sanguigno" [36], were identical at the *Noemi* locus, where both alleles were disrupted by the insertion of different retrotransposons within the sixth exon. One of them, Tcs4, is intact and putatively active; the other one, Tcs6x, is rearranged with two LTRs in tandem in the middle of the insertion (Figure 3A). Both elements are Copia-like LTR retrotransposons, similar to Tcl5 in Limetta dolce and to three elements inserted upstream of *Ruby* in blood orange varieties and in Citron [1, -15] (Figure S4D; Data S1). Consequently, this family of retroelements may be responsible for a large proportion of the

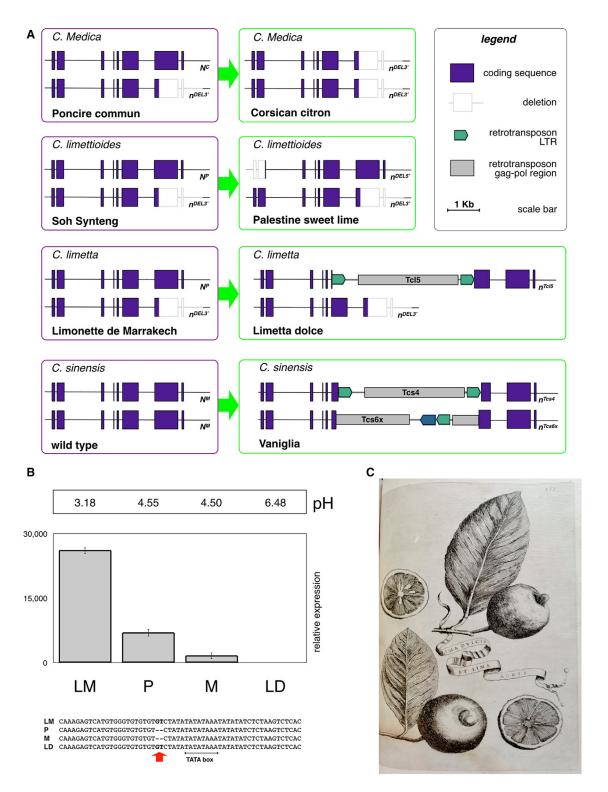


Figure 3. Noemi Is Essential for the Biosynthesis of Flavonoid Pigments and Is a Major Determinant of Fruit Acidity

(A) Allelic constitution of Noemi in wild-type (boxed in purple) and acidless (boxed in green) citrus varieties. Molecular events that resulted in the acidless phenotypes are indicated by a green arrow.

(B) pH of juice, expression of Noemi in fruit and sequences of the promoters of four accessions of C. limetta. LM, Limonette de Marrakech; P, Pomona; M, Millsweet; LD, Limetta dolce. Error bars represent SE. The arrow indicates a 2-bp deletion polymorphism upstream of the TATA box.

phenotypic variation in citrus, particularly in the progeny of interspecific crosses which may carry more active Copia-like retroelements as a result of genome shock [37, 38].

The identification of five independent mutations associated with the acidless phenotype demonstrated that *Noemi* is essential for both the biosynthesis of flavonoid pigments and for fruit acidity and suggests that this regulatory gene may be, in part, responsible for the large spectrum of acidity in commercially grown citrus [14]. In sweet orange, a blood variety with very low acidity, "Tarocco Ferreri," showed a 92% reduction in *Noemi* expression compared to the acidic variety from which it derived, "Tarocco comune" (Figure S4C). Tarocco Ferreri is not an acidless accession since it is able to produce flavonoid pigments and does not contain mutations in the coding sequence of *Noemi*, but the low expression suggests it carries a weak *Noemi* allele or a mutation in a gene controlling *Noemi* expression.

A more direct indication that Noemi is a major determinant of fruit acidity was provided by the analysis of C. limetta. As an apomictic hybrid, its complex genetic constitution (which combines all the three citrus primary species, Figure 1) is fixed within the population. Given that the origin of the hybrid is relatively recent, a single polymorphism in a candidate gene is significant. Besides the fully acidic form and the true acidless variety, we identified two additional accessions producing flavonoid pigments but intermediate levels of fruit acidity (Table 1; Figure S3E). The sequences of the intact Noemi gene in these accessions, named "Pomona" and "Millsweet" [4, 39], were identical to the sequence of the acidic Limonette de Marrakech; the sequences of the regions 1 kb upstream were also identical except for a 2-bp deletion, adjacent to a canonical TATA box in the core promoter, present in both accessions with intermediate acidity (Figures 3B and S3F). The deletion was associated with reduced Noemi expression, which correlated with fruit acidity (Figure 3B).

Our study has identified *Noemi* encoding the bHLH protein that interacts with Ruby to control anthocyanin production in citrus. In this context, it should be possible to use *Noemi* to increase the accumulation of anthocyanins induced by *Ruby* [15]. *Noemi* also controls proanthocyanidin biosynthesis in seeds and is essential for the regulation of fruit acidity. Identification of *Noemi* alleles in Citron allowed us to link modern Mediterranean Citron varieties to those of ancient China, quite possibly facilitated by the adoption of Citron in Jewish culture as an important religious symbol. Although detection of acidless varieties of citrus in the fossil record based on morphological characters of macro-remains is impossible, it might be possible to perform a DNA analysis to identify *Noemi* alleles in the fossil remains from the Mediterranean and to demonstrate their links to citrus of Asian origins.

Noemi is an important determinant of natural variation in fruit acidity in citrus, as evidenced by its alleles in interspecific hybrids like Palestine sweet lime and limetta as well as acidless oranges. This natural variation is associated with recent activity of a family of Copia-like retroelements, particularly evident in inter-

specific hybrids of citrus that may represent a classic example of genome shock as proposed by Barbara McClintock in her Nobel lecture of 1983 [37]. Citrus breeders appear to have used the outcomes of genome shock to select for new variants despite their limited capacity for improvement using conventional breeding.

STAR*METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Plant Material
 - Isolation of Noemi alleles
 - Segregation Analysis
 - O Self-Pollination of 'Poncire commun' Citron
 - O DNA Gel Blot
 - Expression Analysis of Noemi
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, one table, and two data files and can be found with this article online at https://doi.org/10.1016/j.cub. 2018.11.040.

ACKNOWLEDGMENTS

E.B. and C.M. were supported by the Institute Strategic Programs: "Understanding and Exploiting Plant and Microbial Secondary Metabolism" (BB/ J004596/1) and "Molecules from Nature" (BB/P012523/1) from the Biotechnological and Biological Scientific Research Council (BBSRC). We thank Robert Krueger and Manjunath Keremane (USDA-ARS National Clonal Germplasm Repository for Citrus and Dates, Riverside, California, USA) for providing DNA samples and for helpful discussion during the study. We also thank Michel Bachès (Pépinières Bachès, Eus, France) for providing leaves of the Citron accessions listed in Table S1 and information on "Assads Moroccan" Citron. We thank Eliezer Goldschmidt for his encouraging comments on the manuscript.

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 2, 2018 Revised: November 7, 2018 Accepted: November 13, 2018 Published: December 20, 2018

⁽C) Engraving of citrus fruit from a rare copy of "Hesperides" by Ferrari (1646) kept at CREA, Acireale; C. limetta can be recognized by the distinctive prominent nipple. The ribbon is labeled in Latin with the name "Lima Dulcis et Lima Acris" to indicate acidless and acidic fruit, which are morphologically undistinguishable.

REFERENCES

- 1. Butelli, E., Garcia-Lor, A., Licciardello, C., Las Casas, G., Hill, L., Recupero, G.R., Keremane, M.L., Ramadugu, C., Krueger, R., Xu, Q., et al. (2017). Changes in anthocyanin production during domestication of Citrus. Plant Physiol. 173, 2225-2242.
- 2. Ferrari, G.B. (1646). Hesperides sive de malorum aureorum cultura et usu. Volume Libri Quatuor (Romae: Sumptibus Hermanni Scheus).
- 3. Chapot, H. (1950). Un curieux Cédrat marocain (Citrus medica Linné). J. Agric. Tradit. Bot. Appl. 335-336, 506-514.
- 4. Hodgson, R. (1967). History, world distribution, botany and varieties. In The Citrus Industry, W. Reuther, H. Webber, and L. Batchelor, eds. (University of California Press), pp. 431-591.
- 5. Karp, D., and Hu, X. (2018). The Citron (Citrus medica L.) in China. Horticultural Reviews, Volume 45 (Wiley), pp. 143-196.
- 6. Zohary, D., Hopf, M., and Weiss, E. (2012). Domestication of Plants in the Old World, Fourth edition (Oxford University Press).
- 7. Swingle, W., and Reece, P. (1967). The botany of Citrus and its wild relatives. In The Citrus Industry, W. Reuther, H. Webber, and L. Batchelor, eds. (University of California Press), pp. 190-430.
- 8. Fuller, D., Castillo, C., Kingwell-Banham, E., Qin, L., and Weisskopf, A. (2017). Charred pummelo peel, historical linguistics and other tree crops: Approaches to framing the historical context of early Citrus cultivation in East, South and Southeast Asia. In AGRUMED: Archaeology and history of citrus fruit in the Mediterranean, V. Zech-Matterne, and G. Fiorentino, eds. (Publications du Centre Jean Bérard), pp. 29-48.
- 9. Lotsy, J. (1925). Species or linneon? Genetica 7, 487-506.
- 10. Tanaka, T. (1977). Fundamental discussion of citrus classification. Studia Citrologica 14, 1-6.
- 11. Scora, R.W. (1975). Symposium on biochemical systematics, genetics and origin of cultivated plants 0.9. history and origin of citrus. B. Torrey Bot. Club. 102, 369-375.
- 12. Barrett, H., and Rhodes, A. (1976). A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. Syst. Bot. 1, 105-136.
- 13. Curk, F., Ollitrault, F., Garcia-Lor, A., Luro, F., Navarro, L., and Ollitrault, P. (2016). Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. Ann. Bot. 117, 565-583.
- 14. Wu, G.A., Terol, J., Ibanez, V., López-García, A., Pérez-Román, E., Borredá, C., Domingo, C., Tadeo, F.R., Carbonell-Caballero, J., Alonso, R., et al. (2018). Genomics of the origin and evolution of Citrus. Nature 554. 311-316.
- 15. Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., and Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell 24, 1242-1255.
- 16. Ramsay, N.A., and Glover, B.J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 10, 63-70.
- 17. Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. 20, 176-185.
- 18. Spelt, C., Quattrocchio, F., Mol, J., and Koes, R. (2002). ANTHOCYANIN1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. Plant Cell 14, 2121-2135.
- 19. Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J., and Koes, R. (2006). PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. Plant Cell 18, 1274-1291.
- 20. Faraco, M., Spelt, C., Bliek, M., Verweij, W., Hoshino, A., Espen, L., Prinsi, B., Jaarsma, R., Tarhan, E., de Boer, A.H., et al. (2014). Hyperacidification

- of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color. Cell Rep. 6, 32-43.
- 21. Luro, F., Venturini, N., Costantino, G., Paolini, J., Ollitrault, P., and Costa, J. (2012). Genetic and chemical diversity of citron (Citrus medica L.) based on nuclear and cytoplasmic markers and leaf essential oil composition. Phytochemistry 77, 186-196.
- 22. Cultrone, A., Cotroneo, P.S., and Recupero, G.R. (2010). Cloning and molecular characterization of R2R3-MYB and bHLH-MYC transcription factors from Citrus sinensis. Tree Genet, Genomes 6, 101-112.
- 23. Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., and Lepiniec, L. (2000). The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. Plant Cell 12, 1863-1878.
- 24. Wen, J., Li, Y., Qi, T., Gao, H., Liu, B., Zhang, M., Huang, H., and Song, S. (2018). The C-terminal domains of Arabidopsis GL3/EGL3/TT8 interact with JAZ proteins and mediate dimeric interactions. Plant Signal. Behav.
- 25. Ramadugu, C., Keremane, M.L., Hu, X.L., Karp, D., Federici, C.T., Kahn, T., Roose, M.L., and Lee, R.F. (2015). Genetic analysis of citron (Citrus medica L.) using simple sequence repeats and single nucleotide polymorphisms. Sci. Hortic. (Amsterdam) 195, 124-137.
- 26. Nicolosi, E., La Malfa, S., El-Otmani, M., Negbi, M., and Goldschmidt, E.E. (2005). The search for the authentic citron (Citrus medica L.): Historic and genetic analysis. HortScience 40, 1963-1968.
- 27. Goldschmidt, E.E. (2017). New insights in citron (Citrus medica L.) genomics and fruit development. HortScience 52, 823-826.
- 28. Isaac, E. (1959). Influence of religion on the spread of citrus: The religious practices of the Jews helped effect the introduction of citrus to Mediterranean lands. Science 129, 179-186.
- 29. Chapot, H. (1962). Trois citrus Marocains. Al Awamia [Rabat] 2, 47-81.
- 30. Zhou, J.M., Hirata, Y., Nou, I.S., Shiotani, H., and Ito, T. (2002). Interactions between different genotypic tissues in citrus graft chimeras. Euphytica 126, 355-364.
- 31. Frank, M.H., and Chitwood, D.H. (2016). Plant chimeras: The good, the bad, and the "Bizzaria". Dev. Biol. 419, 41-53.
- 32. Inzenga, G. (1915). Agrumi siciliani. Monografia. Annali della R. Stazione sperimentale di agrumicoltura e frutticoltura di Acireale, Volume III (Acireale), pp. 1-42.
- 33. Aprile, A., Federici, C., Close, T.J., De Bellis, L., Cattivelli, L., and Roose, M.L. (2011). Expression of the H+-ATPase AHA10 proton pump is associated with citric acid accumulation in lemon juice sac cells. Funct. Integr. Genomics 11, 551-563.
- 34. Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., et al. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. 32, 656-662.
- 35. Chapot, H., and Praloran, J. (1955). Les graines de Citrus. In XIV International Horticulture Congress, Volume 2, J.P. Nieuwstraten, ed. (H. Veenman and Zonen), pp. 1294–1323.
- 36. Huet, R., and Chapot, H. (1964). Oranges et citrons roses. Al Awamia [Rabat] 11, 21-29.
- 37. McClintock, B. (1984). The significance of responses of the genome to challenge. Science 226, 792-801.
- 38. Comai, L., Madlung, A., Josefsson, C., and Tyagi, A. (2003). Do the different parental "heteromes" cause genomic shock in newly formed allopolyploids? Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 1149-1155.
- 39. Ramadugu, C., Razi, M.F., Keremane, M.L., Scora, R.W., and Roose, M.L. (2017). The Lime Botany, Production and Uses. Systematic Classification, Distribution and Botany (CABI Publishing).

STAR*METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
4-(Dimethylamino)cinnamaldehyde	Sigma-Aldrich Merck	D4506
Critical Commercial Assays		
DNeasy Plant Mini Kit	QIAGEN	69106
RNeasy Plant Mini Kit	QIAGEN	74904
Phusion High-Fidelity DNA Polymerase	Thermo Fisher	F530L
pGEM®-T Easy Vector Systems	Promega	A1360
SuperScript III Reverse Transcriptase	Thermo Fisher	18080044
High-Capacity cDNA Reverse Transcription Kit	Thermo Fisher	4368814
ABI PRISM 7000 Sequence Detection Systems	Thermo Fisher	4328895
SYBR Green PCR Master Mix	Thermo Fisher	4309155
Custom DNA sequencing	Eurofins	https://www.eurofinsgenomics.eu/
Deposited Data		
Nucleotide sequence of Noemi in citrus accessions	This study	GenBank: MK139964-MK139972
Experimental Models: Organisms/Strains		
List of citrus accessions presented in Tables 1 and S1	This study	N/A
Oligonucleotides		
List of oligonucleotides presented in Table S2	This study	N/A
Software and Algorithms		
Citrus Genome Assembly and Annotation	Citrus clementina v1.0	https://phytozome.jgi.doe.gov/
Citrus Genome Assembly and Annotation	Citrus sinensis v1.1	https://phytozome.jgi.doe.gov/
Citrus Genome Assembly and Annotation	Citrus sinensis Annotation project	http://citrus.hzau.edu.cn/
Sequence alignment	MultAlin	http://multalin.toulouse.inra.fr/multalin/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Eugenio Butelli (eugenio.butelli@jic.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Leaves and fruit were of citrus accessions were obtained from the sources listed in Table 1 and Table S1. Plant material was analyzed as described in Method Details.

METHOD DETAILS

Plant Material

Citrus leaves were ground in liquid nitrogen and DNA and RNA were extracted using the DNeasy Plant Mini Kit and the RNeasy Plant Mini Kit (QIAGEN) according to the manufacturer's instructions. Fruit juice was obtained using a conventional citrus squeezer and pH was measured in three replicates with a standard combination Ag/AgCl pH electrode. The presence of anthocyanins in young leaves and flowers was determined by visual inspection and confirmed measuring the absorbance at 530 nM of acidified methanol extracts. The presence of proanthocyanidins in seeds was determined by visual inspection and confirmed by staining for 30 min with 0.3% (w/v) DMACA (p-Dimethylaminocinnamaldehyde, Sigma) in methanol and 6 M HCl (1:1, v/v) followed by several washing steps with 70% ethanol.

Isolation of Noemi alleles

Full-length *Noemi* alleles were isolated by PCR using primers EB-072 and EB-076 designed within the 5'UTR and 3'UTR regions, respectively. PCR fragments were used directly for sequencing or cloned into the pGEM-T Easy vector (Promega) when two alleles



having the same size were present. The alleles containing retrotransposons where isolated using primers EB-088 and EB-076. Appropriate primers were used to obtain the complete sequences and confirm the presence of the retrotransposons. The alleles with a deletion in the 3' region, n^{DEL3'}, where identified using primers EB-093 and EB-94 followed by sequencing. The alleles with a deletion in the 5' region, $n^{DELS'}$, and the core promoter regions where isolated using primers EB-111 and EB-076. The determination of the intron - exon structure was performed after isolation of full-length cDNA clones from leaves of wild-type Citron and sweet orange. Total RNA was retrotranscribed using Superscript III reverse transcriptase (Thermo Fisher) and amplified using primers EB-105 and EB-106 in Citron followed by direct sequencing or, for sweet orange, with primers EB-115 and EB-116, equipped with attB sites, followed by cloning into pDONR 207 (Invitrogen). Primer sequences are listed in Table S2. The sequences of the different Noemi alleles are provided in Data S2 and have been deposited in GenBank under the accession numbers MK139964 to MK139972.

Segregation Analysis

Seeds were extracted from three fruits of 'Poncire commun' Citron, washed with water and sterilized with 10% bleach for three hours with shaking. Germination of seeds was carried out in tissue culture conditions placing seeds in Phytatray II vessels (Sigma) containing MS medium with 0.8% agar; seeds were kept at 23°C with 16 h of light and 8 h of dark. One tray, accidentally contaminated by the common mold Cladosporium (as determined by sequencing), showed unequivocal production of anthocyanins in five out of eight seedlings. All the eight seedlings were used for PCR analysis using primers EB-093 and EB-094.

Self-Pollination of 'Poncire commun' Citron

Self-pollination was realized between April 23rd and May 20th 2013 at the INRA-CIRAD Citrus germplam of San Giuliano, Corsica, France. One tree of 'Poncire commun' was covered with insect proof net. Fruits were harvested on December 9th and seeds were extracted and sown on the same day. Twenty plantlets were grafted on Citrus volkameriana rootstock on March 17th 2014. The first flowers were obtained during the spring 2015. Three trees (two with purple and one with white flowers) are conserved in a greenhouse in San Giuliano. Every year, the tree with white flowers produces acidless fruit and trees with purple flowers acidic fruit. Mature leaves from each phenotype were used for PCR analysis using primers EB-093 and EB-094.

DNA Gel Blot

Leaves of different varieties of lemon were ground in liquid nitrogen and DNA was extracted using caesium chloride density gradient purification. DNA (10 µg per sample) was digested with HindIII restriction enzyme for 5 h and then separated by electrophoresis. Denatured DNA was transferred to nitrocellulose membrane filters. Filters were hybridized with randomly primed ³²P-labeled probe of the Ruby gene overnight at 60°C and washed in 0.1 × SSC, 0.5% SDS at 60°C for 2 h before exposure to X-ray film (Fuji RX-100).

Expression Analysis of Noemi

Quantification of Noemi expression in fruit of different varieties of limetta and sweet orange was performed by qRT-PCR. Total RNA was extracted from 3 mL of juice, previously filtered under sterile conditions. One volume of extraction buffer (0.2 M Tris-HCl pH 8, 50 mM EDTA, 0.2 M NaCl, 2% (w/v) SDS), one volume of phenol and 60 μl of β-mercaptoethanol were added to 3 mL of juice. After an incubation at 50°C for 5 minutes, samples were centrifuged at 4000 rpm for 15 minutes. The agueous phase was extracted for two times with one volume of chloroform:isoamyl alcohol (24:1, v/v) after centrifugations at 4000 rpm for 15 minutes each. Half volume of 6 M LiCl was added to the new aqueous phase, and RNA was left to precipitate overnight at -20°C. After centrifugation and washing with 70% (v/v) ethanol, RNA was resuspended in RNase-free water. DNase-treated total RNA was further purified using the RNA Cleanup protocol (QIAGEN) and retrotranscribed into cDNA using a High-Capacity cDNA Archive kit (Thermo Fisher). Quantitative real-time PCR was performed in optical 96-well plates with an ABI Prism 7000 sequence detection system (Thermo Fisher). A PCR mixture (final volume 25 μl) containing 15 μL Power SYBR Green mix, 0.2 μM each of gene-specific primers and 100 ng of cDNA sample was prepared using the protocol for Power SYBR Green PCR Master Mix (Thermo Fisher). The following standard thermal profile was used for all PCR reactions: 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 s, and 60°C for 1 min. Three replicates were assayed, and a no-template negative control (water control) was performed. The analyses used the relative quantification standard curve method.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of Noemi expression was performed by gRT-PCR; three replicates were assayed. Error bars in Figures 3 and S4C represent standard error of the mean (SE) and were determined using Excel 2017 (Microsoft, Redmond, WA).

Supplemental Information

Noemi Controls Production of Flavonoid Pigments and Fruit Acidity and Illustrates the

Domestication Routes of Modern Citrus Varieties

Eugenio Butelli, Concetta Licciardello, Chandrika Ramadugu, Marie Durand-Hulak, Alessandra Celant, Giuseppe Reforgiato Recupero, Yann Froelicher, and Cathie Martin

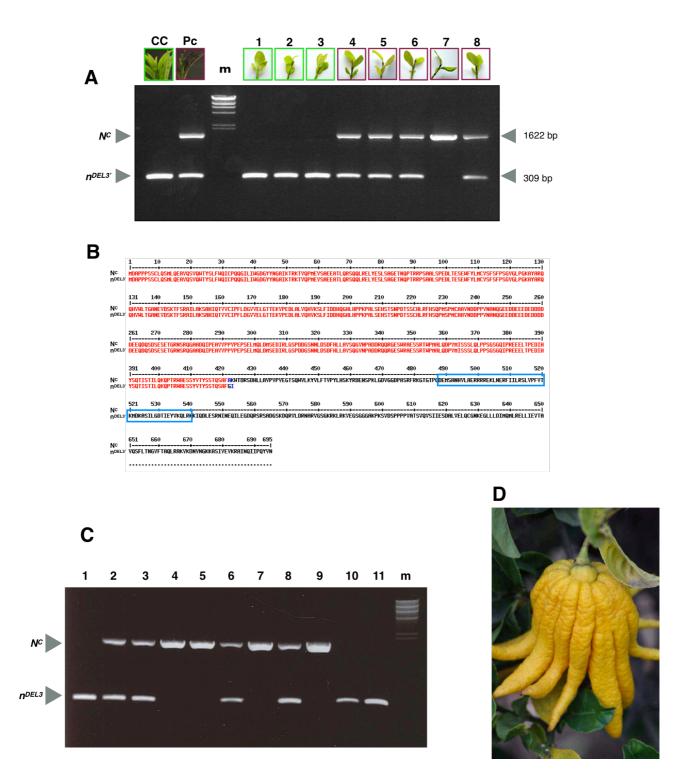


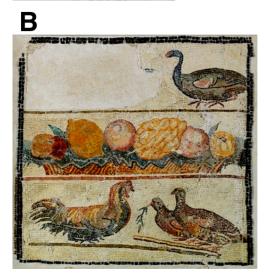
Figure S1. Genetic characterization of Citron accessions Related to Figures 2 and 3, Table 1 and Table S1.

- (A) PCR analysis using genomic DNA extracted from leaves of 'Corsican' Citron (CC), 'Poncire commun' (Pc) and from seedlings grown from seeds of 'Poncire commun' showing absence (1 to 3) or presence (4 to 8) of anthocyanin pigmentation.
- (B) Sequence alignment of the proteins encoded by the *Noemi* alleles N^C and $n^{DEL3'}$. The predicted bHLH domain is boxed in blue.

- (C) PCR analysis showing the presence of the non-functional alleles $n^{DEL3'}$ in Citron accession of different origin (1, 'Assads Moroccan'; 2, 'Yemen'; 3, 'Greek'; 4, 'Florentine'; 5, 'Diamante'; 6, 'Buddha's Hand USA'; 7, 'Buddha's Hand Taiwan'; 8, 'Buddha's Hand Japan'; 9, 'Buddha's Hand variegated'; 10, 'Buddha's Hand Qingpi'; 11, 'Buddha's Hand Aihua').
- (D) Fruit of 'Buddah's hand USA' citron (*C. medica* var. *sarcodactylis*) with its peculiar shape segmented into finger-like sections. Image from 'Citrus ID Tools' (http://idtools.org/id/citrus/citrusid/).







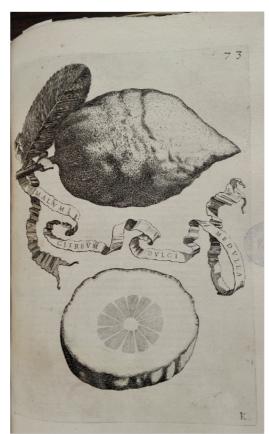


Figure S2. Citron fruit portrayed in historical mosaics and engravings Related to Figure 2 and Table 1.

- (A) Citron, house of the Dionysian Procession, El-Jem, Tunisia, mosaic. Image from akgimages / Gilles Mermet.
- (B) Citron fruit represented in the centre of a mosaic displayed at the National Roman Museum, Palazzo Massimo alle Terme, Rome. A lemon is also illustrated. Image from Open Edition Books (https://books.openedition.org/pcjb/docannexe/image/2194/img-6.jpg).
- (C) Engraving of Citron fruit from a rare copy of 'Hesperides' by Ferrari (1646) kept at CREA, Acireale; the fruit is described in Latin as '*Malum Citreum Dulci Medulla*' (Citron with sweet pulp).
- (D) Citrus fruit under the Medici family in Florence. 'Melangoli, Cedri e Limoni' by Bartolomeo Bimbi, 1715. An extensive collection of citrus fruit is depicted with great accuracy. The fruit at the centre of the canvas (number 14 and circled in red) is named as 'Melangola', a vernacular term used to define a 'Limetta dolce' in the XVIII century. Image from Wikimedia

(https://commons.wikimedia.org/wiki/File:Bartolomeo_bimbi,_melagoli,_cedri_e_limoni,_17 15,_01.JPG).

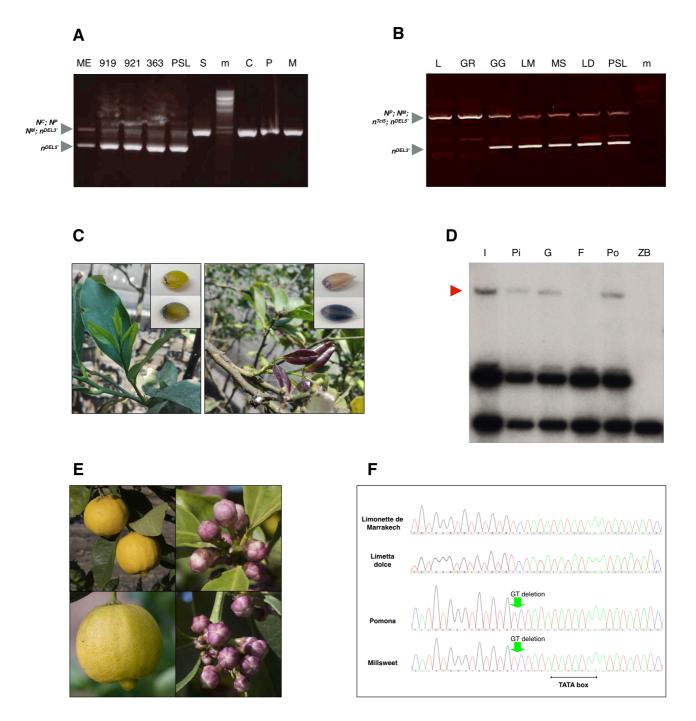


Figure S3. Noemi alleles in in Citron-derived hybrids Related to Figure 3 and Table 1.

(A) PCR analysis showing the presence of the non-functional alleles $n^{DEL5'}$ exclusively in in the acidless accessions of *C. limettioides* (ME, 'Mary Ellen'; 919, unnamed-CRC919; 921, unnamed-CRC921; 363, unnamed-CRC363; PLS, 'Palestine sweet lime') and not in the acidic form (S, 'Soh Synteng') or in the primary species (C, 'Diamante' Citron; P, 'Chandler' Pummelo; M, 'Ponkan' Mandarin) involved in the origin of *C. limettioides*.

(B) PCR analysis showing the presence of the non-functional alleles $n^{DEL3'}$ in *C. limetta* (LM, 'Limonette de Marrakech'; MS, 'Millsweet'; LD, 'Limetta dolce') and *C. limettioides*

- (PLS, 'Palestine sweet lime'). Wild-type lemon (L) does not contain $n^{DEL3'}$ which is present in green leaves of the sweet lemon 'Girotta' (GG) but not in red leaves (GR), indicating that 'sweet lemons' are chimeric varieties involving *C. limetta*.
- (C) A sweet lemon accession, 'Poros', displaying chimeric green leaves (left panel) and revertant wild-type pigmented leaves (right panel) on the same plant. The revertant branches produce acidic fruit and dark seeds with proanthocyanidins; the inset figures show one seed from the two types of fruit before and after staining with DMACA reagent. (D) Southern-blot analysis of genomic DNA from different lemon varieties digested with *HindIII* and probed with a ³²P-labeled probe of the *Ruby* gene. The red triangle indicates hybridizations bands that are only present in sweet lemon accessions (I, 'ISA'; Pi, 'Pispisa'; G, 'Girotta'; Po; 'Poros') but not in true lemons (F, 'Femminiello'; ZB, 'Zagara Bianca'). These bands reflect the presence of an additional allele derived from *C. limetta* and indicate that sweet lemons are periclinal graft chimeras. The diagnostic bands are fainter that those only present in true lemons because of the low DNA contribution of the L1 layer and because a mixture of green (chimeric) and red (revertant wild-type) leaves growing on the same plants was used.
- (E) Fruit and flower buds of 'Pomona' (upper panels) and 'Millsweet' (lower panels).
- (F) DNA sequence chromatograms showing a GT dinucleotide deletion affecting the promoter of the active Noemi allele in two accessions of C. limetta, 'Pomona' and 'Millsweet', with reduced fruit acidity compared to the wild-type acidic form, 'Limonette de Marrakech'. The promoter in the fully acidless form, 'Limetta dolce', is also wild-type but the coding sequence has been disrupted by the insertion of a retrotransposon. The second Noemi allele, n^{DEL3} ', is non-functional and identical in all the four accessions.

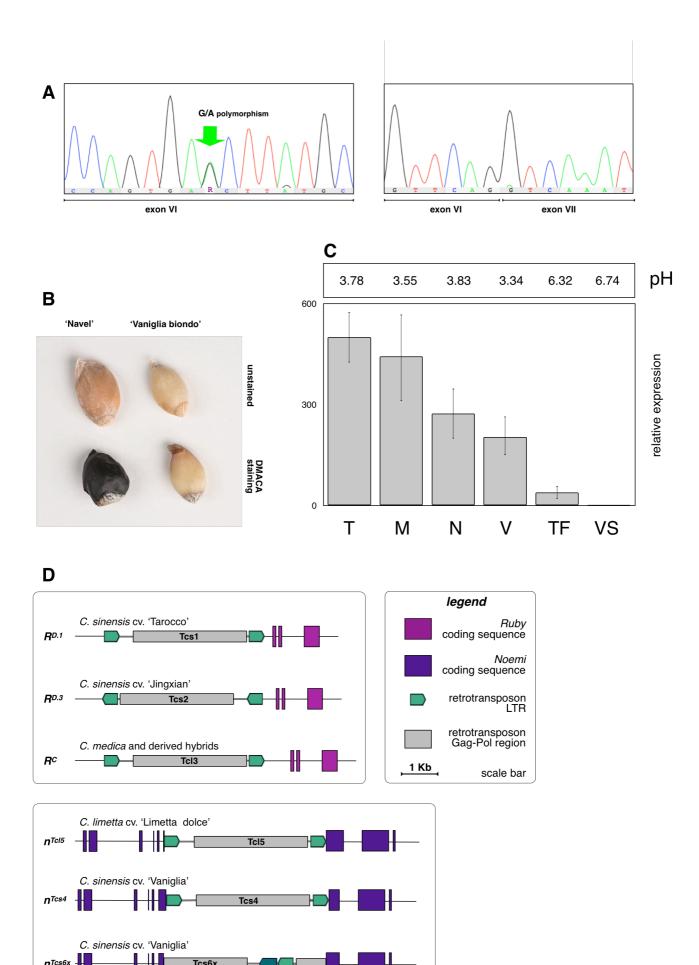


Figure S4. *Noemi* in different accessions of sweet orange Related to Figure 3 and Table 1.

- (A) Sequence chromatogram showing biallelic expression of *Noemi* in fruit of wild-type ('Navel') sweet orange. The green arrow on the left panel indicates one of the three polymorphisms identified in the coding region. On the same chromatogram, the region at the junction between two exons is also shown (right panel) to confirm that the sequence corresponds to the cDNA of *Noemi* and not to genomic DNA. *Noemi* expression was not detected in the acidless ('Vaniglia sanguigno') accession.
- (B) Seed of wild-type ('Navel') and acidless ('Vaniglia biondo') sweet orange accessions showing a different colouration in the inner seed coat. The use of DMACA staining indicates that the lighter colour of the seeds in the acidless variety is due to lack of production of proanthocyanidins.
- (C) Measurement of pH in juice and expression of *Noemi* in sweet orange accessions with different levels of acidity. T, 'Tarocco comune'; M, 'Moro'; N, 'Navel'; V, 'Valencia'; TF, 'Tarocco Ferreri'; VS, 'Vaniglia sanguigno'. Error bars show SE of the mean.
- (D) Schematic diagram showing the insertion of six LTR retrotransposons at the *Ruby* (upper panel) or *Noemi* (lower panel) loci. The names of the corresponding alleles and the citrus accessions where the LTR retrotransposons have been identified are indicated.

Table S1. Allelic constitution of *Noemi* in the accession of Citron used in this study Related to Figures 2, 3 and Table 1.

Common name	Noemi alleles		Source	Accession
Poncire commun	N C	n ^{DEL3'}	D	SRA701
Corsican	n ^{DEL3'}	n ^{DEL3'}	D	SRA613
Assads Moroccan	n ^{DEL3'}	n ^{DEL3'}	E	n.a.
Yemen	N C	n ^{DEL3'}	E	n.a.
Greek	N C	n ^{DEL3'}	E	n.a.
Florentine	N C	Nc	E	n.a.
Diamante	N C	N c	В	Palazzelli Certif.
Buddha's Hand - USA	N C	n ^{DEL3'}	С	CRC3768
Buddha's Hand - Taiwan	N C	Nc	E	n.a.
Buddha's Hand - Japan	N C	n ^{DEL3'}	E	n.a.
Buddha's Hand - variegated	N C	Nc	E	n.a.
Buddha's Hand - Qingpi	n ^{DEL3'}	n ^{DEL3'}	С	CM1
Buddha's Hand - Aihua	n ^{DEL3'}	n ^{DEL3'}	С	CM2
Buddha's Hand - Spain	Nc	N c	E	n.a.
Buddha's Hand - China	N c	N C	Е	n.a.

Wild-type alleles of *Noemi* are highlighted in purple; mutated alleles are highlighted in green. Citrus accessions were obtained from the following sources: B: CREA, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia agraria, Olivicoltura Frutticoltura Agrumicoltura, 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, Pépinières Bachès, Eus, France.

n.a.: not available.

Table S2. Sequences of the primers used in the study

Name	Sequence	Description	
EB-072	GCGGTGAATACAGTCGGTAAAGGA	Amplification of full- length <i>Noemi</i>	
EB-076	ACCCAGCTATCCAAATTATTACAATTCGT		
EB-093	GCTGCAGGTCAAATTCCACGGGA	Identification of <i>n</i> ^{DEL3'}	
EB-094	GCCATGCTATACTCTGCTGACCGA		
EB-111	ACTCCCTAGCCTCTCATTTCTTCCTTCCA	Identification of <i>n</i> ^{DEL5'}	
EB-114	GACAGTTGACAGGAAGAGAGAGAACA	and promoter analysis	
LD-114	anonatianonaanananananan	and promoter analysis	
EB-088	GCAGACTGTGGTATGCATTCCT	Isolation of n^{Tcl5} , n^{Tcs4} ,	
EB-125	CTGCTACCTATCATAACCGGTTCCTGT	n ^{Tcs6x}	
EB-169	CAGGAAAGGAGAACTTGGTTTGCAGGT		
EB-455	GGGTTCACTCCTTGGCTAACTGCGA		
EB-464	CCACCGGCCTTTGACATTTCTCCCA		
EB-105	GGAATGGATGCGCCGCCAAGT	Isolation of <i>Noemi</i>	
EB-106	GCTATGAGTGAGTTAATTGACATACTGGGGT	cDNA-from Citron	
LD-100	dotatadatadattaattadatadtadadt	ODIVIT HOITI ORIGIT	
EB-115	GGGGACAAGTTTGTACAAAAAAGCAGGCTAT	Isolation of <i>Noemi</i>	
	GGATGCTCCGCCGCCGAGTA	cDNA-from sweet	
EB-116	GGGGACCACTTTGTACAAGAAAGCTGGGTGA	orange	
	GTGAGTTAATTGACATACTGGGGT		
No ami DT F		Ma ami ayaya asiga iz	
NoemiRT-Fw	CAGGAACCGGTTATGATAGGTAGC	Noemi expression in	
NoemiRT-Rev	TCTGGCGTCAATTCTTCTTCCGGTG	sweet orange qRT-PCR	
NoemiWT-Fw	AGTCCCGCAACATCAACAAG	Noemi expression in	
NoemiWT-Rev	TTTGATCTTTGGAGCCATCC	limetta qRT-PCR	