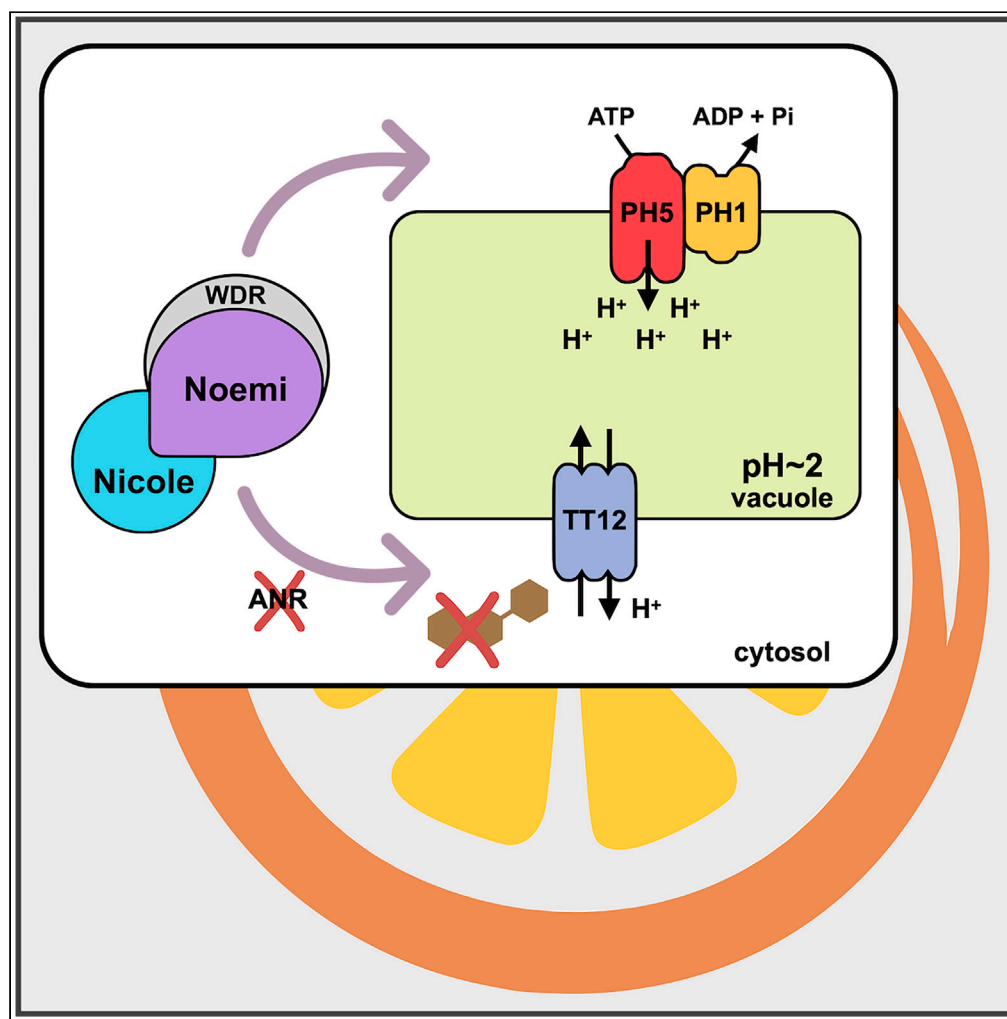**Article**

Distinctive acidity in citrus fruit is linked to loss of proanthocyanidin biosynthesis

Elliott Atkins,
Emanuele Scialò,
Chiara Catalano,
..., Andrés Garcia-Lor, Cathie Martin,
Eugenio Butelli

cathie.martin@jic.ac.uk (C.M.)
eugenio.butelli@jic.ac.uk (E.B.)

Highlights

*Nicole i*s essential for the regulation of fruit acidity in citrus

In a natural low-acid mutant of sweet orange, *Nicole* is interrupted by a retrotransposon

Nicole upregulates vacuolar proton pumps and several proanthocyanidin biosynthetic genes

Incomplete activation of the proanthocyanidin pathway may explain extreme acidity in citrus

Article

Distinctive acidity in citrus fruit is linked to loss of proanthocyanidin biosynthesis

Elliott Atkins,[1,5,8] Emanuele Scialò,[1,6,8] Chiara Catalano,[1,6] Carmen Caballero Hernández,[1,7] Eva Wegel,[1] Lionel Hill,[1] Concetta Licciardello,[2] Leandro Peña,[3] Andrés Garcia-Lor,[4] Cathie Martin,[1,*] and Eugenio Butelli[1,9,*]

SUMMARY

The distinctive acidity of citrus fruit is determined by a regulatory complex of MYB and bHLH transcription factors together with a WDR protein (MBW complex) which operates in the unique juice vesicles of the fruit. We describe a mutation affecting the MYB protein, named Nicole, in sweet orange and identify its target genes that determine hyperacidification, specifically. We propose that the acidity, typical of citrus fruits, was the result of a loss of the ability of Nicole to activate the gene encoding anthocyanidin reductase, an enzyme essential for the synthesis of proanthocyanidins, which are absent in citrus fruit.

INTRODUCTION

Citrus fruits are well known for their acidity and, with a pH of around 2, lemon is the most acidic common constituent of the human diet. Many wild and cultivated citrus species and hybrids (including citron, most limes, "true" mandarins, trifoliate, and sour orange) share similarly high levels of acidity. The acidity of citrus fruit may have underpinned early uses as medicine and antiseptics. However, thousands of years of domestication have resulted in accessions with moderate acidity with a typical pH of around 3.5 for some mandarins and sweet orange fruit.

Together with total soluble sugars, acidity is the major determinant of taste and quality of citrus fruit. The very low pH reflects the content of citric acid (citrate), which is produced within the juice vesicles, the unique structures that form the edible part of citrus fruit.[1,2] At the cellular level, citrate biosynthesis requires the coordinated activity of many enzymes in different compartments. It is generally assumed that reduced acidity was a key trait for citrus domestication, and genetic differences in citrate biosynthetic genes may explain differences in acidity between wild and cultivated citrus and between lemon and orange.[3,4]

Besides biosynthesis and degradation, a key step in citrate metabolism is its transport to the vacuole. The current understanding is that this process is achieved through a mechanism referred to as the "acid trap," dependent on the pH of the vacuole and the electrochemical H^+ gradient across the tonoplast.[5,6] Acidification of the vacuole can be achieved through the activity of vacuolar pyrophosphatases and vacuolar ATPases, proton pumps ubiquitous in plants. In some species including citrus, however, a second distinct mechanism has been identified which requires the activity of specific P-type ATPases involved in proton transport. These P-type ATPases are essential for hyperacidification, driving the pH to values considerably lower than in a typical plant vacuole (pH of around 5.5). In petunia, two P-type ATPases (named PH5 and PH1) interact functionally to induce vacuolar hyperacidification and mutants were identified by variation in pH-dependent flower color.[7] The process of hyperacidification is controlled by a transcriptional activation complex formed by an MYB (M) transcription factor, a bHLH (B) transcription factor, and a WD-repeat (W) protein, also involved in regulating proanthocyanin and anthocyanin biosynthesis.[8,9] Despite their broad evolutionary distance, a similar mechanism appears to control pH in citrus fruit, where the expression of *PH5* correlates with the level of acidity.[10,11] Evidence for a conserved regulatory mechanism emerged with the genetic characterization of the "acidless" phenotype,[12,13] observed in a peculiar group of citrus mutants where the complete lack of anthocyanins in young leaves and flowers is accompanied by the absence of proanthocyanidins (PAs; also known as condensed tannins) in seeds and, most notably, with an extreme reduction in fruit acidity (involving about two- or three-unit changes in pH compared to the corresponding wild-type accessions). These distinct physiological processes are controlled by a gene with pleiotropic functions encoding a bHLH transcription factor (named Noemi or CitAN1), leading to the hypothesis that the same bHLH and WD-repeat proteins can form three distinct MBW complexes with distinct MYB transcription factors which provide the DNA-binding specificity that drives the MBW complex to the activation of specific promoters and, therefore, to the control of specific processes (Figure 1A).[14] Hence, while mutations in the bHLH *Noemi* gene affect all the three processes,[12,13] mutations in a certain

[1]John Innes Centre, Norwich NR4 7UH, UK
[2]CREA, Research Center for Olive Fruit and Citrus Crops, Corso Savoia 190, 95024 Acireale, Italy
[3]Instituto de Biologia Molecular y Celular de Plantas – Consejo Superior de Investigaciones Científicas, Universidad Politécnica de Valencia, Valencia, Spain
[4]Centro de Citricultura y Producción Vegetal, Instituto Valenciano de Investigaciones Agrarias, Valencia, Spain
[5]Present address: Tropic Biosciences Ltd, Norwich Research Park Innovation Center, Norwich NR4 7GJ, United Kingdom
[6]Present address: Department of Agriculture, Food and Environment (Di3A), University of Catania, Via Santa Sofia 100, 95123 Catania, Italy
[7]Present address: Moyca Grapes S.L. Pol. Ind. El Saladar, C. Arquímedes, s/n, 30850, Murcia, Spain
[8]These authors contributed equally
[9]Lead contact
*Correspondence: cathie.martin@jic.ac.uk (C.M.), eugenio.butelli@jic.ac.uk (E.B.)
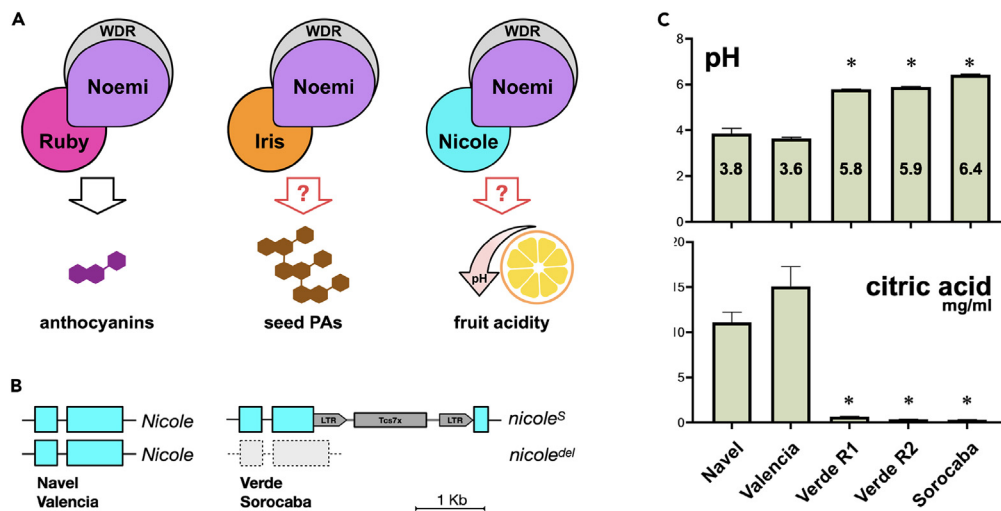https://doi.org/10.1016/j.isci.2024.110923

Figure 1. Mutations in *Nicole* are associated with very low fruit acidity and dramatic reduction in citric acid content of sweet oranges

(A) Working hypothesis showing how the same bHLH (Noemi) and WDR proteins can form three distinct MBW complexes with different MYB transcription factors (Ruby, Iris, and Nicole), driving the expression of different sets of genes and regulating different physiological processes. The red question marks indicate the previously unconfirmed identities of the MYB transcription factors controlling PA accumulation in seeds or fruit acidity.

(B) Allelic constitution of *Nicole* in sweet orange varieties "Navel" and "Valencia" (both wild type for acidity) or "Lima Verde" and "Lima Sorocaba".

(C) Measurement of pH and citric acid content in juice obtained from individual fruits of different varieties of sweet orange. Error bars represent the standard deviation of the mean ($n \geq 3$). Asterisks indicate comparative statistical significance relative to "Navel": * $p < 0.0001$.

See also Figures S1 and S2 and Data S1.

MYB genes would only have consequences on one of the three specific phenotypic traits. While the MYB transcription factor, Ruby, has been unequivocally established as the MYB transcription factor that activates anthocyanin biosynthesis in citrus,[15–17] the identity of the genes encoding the other two MYB proteins remains uncertain. The aim of this study was the identification and characterization of those MYB transcription factors controlling PA production in seeds and fruit hyperacidification (Figure 1A).

RESULTS

### Isolation of Iris, MYB regulator of PA biosynthesis in seeds of sweet orange

A gene encoding a MYB transcription factor regulating PA biosynthesis in seeds, *TT2*, has been extensively characterized in many species including Arabidopsis.[18–20] In citrus, however, the ortholog of *TT2* contains mutations including a G/A transition in the start codon conserved in all primary species (mandarin, pummelo, and citron) according to the citrus pangenome (http://citrus.hzau.edu.cn/). We confirmed the presence of the start codon mutations and stop mutations within the first half of the coding sequence of *TT2* in citrus primary species and in several hybrids, including sweet orange. Then, we analyzed three genes encoding MYB transcription factors which, based on sequence homology and phylogenetic analysis (Figure S1A), appeared related to TT2, clustered within the same subgroup 5a, and therefore had the potential to activate PA biosynthesis. We named them Aris, Marys, and Iris (Table S1). Marys was the most closely related to TT2 and annotated as such in the sweet orange v1.1 Phytozome reference genome (https://phytozome-next.jgi.doe.gov/info/Csinensis_v1_1). Iris was highly similar (64% amino acid identity) to VvMYBPA1, a well-characterized transcription factor that regulates PA accumulation in seeds and developing berries of grape.[21] *Marys* and *Iris* were highly expressed in the inner seed coat isolated from different varieties of sweet orange (Figure S1B). The transcript of *Aris* was barely detectable, but we were able to isolate the corresponding full-length coding sequence from very young developing fruit. To test the functionality of the three genes, we transiently expressed their coding sequences in leaves of *N. benthamiana* and treated the extracts with DMACA (4-dimethylaminocinnamaldehyde) reagent, which specifically stains PAs and their monomers (catechins and epicatechins) a purple-blue color. Only *Iris* was able to induce the accumulation of these flavonoid compounds (Figure S1C). To confirm the functionality, or lack thereof, we used a dual luciferase reporter assay to assess the ability of Aris, Marys, and Iris to activate the promoters of anthocyanidin synthase (*ANS*), a gene essential for PA biosynthesis. Iris strongly activated this promoter, while Aris, Marys showed only very weak, most likely non-specific, activity (Figure S1D). When stably expressed in tobacco under the control of the *CaMV 35S* promoter, Iris was able to induce PA biosynthesis, particularly in flowers (Figure S1E). We concluded that Iris is the MYB controlling seed PAs in citrus (Figure 1A).

Isolation of Nicole, MYB regulator of fruit acidity in sweet orange

By analogy to petunia, a gene encoding a MYB protein with homology to PhPH4 has been considered the regulator of fruit acidity[13,22,23] and, recently, the use of CRISPR-Cas9 technology has resulted in edited lines of *Fortunella hindsii* (Hongkong kumquat), with substantial reduction in fruit acidity and citrate levels.[24] To confirm the identity of the gene encoding the MYB transcription factor controlling hyperacidification of
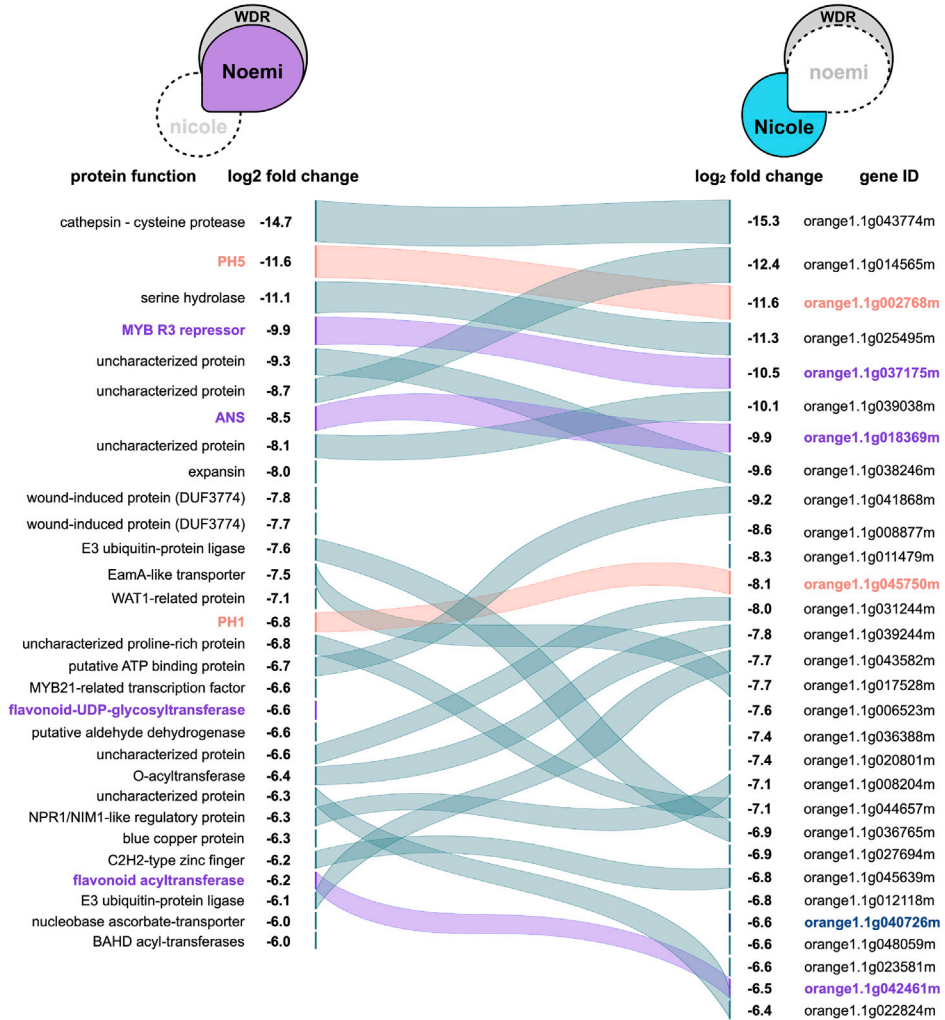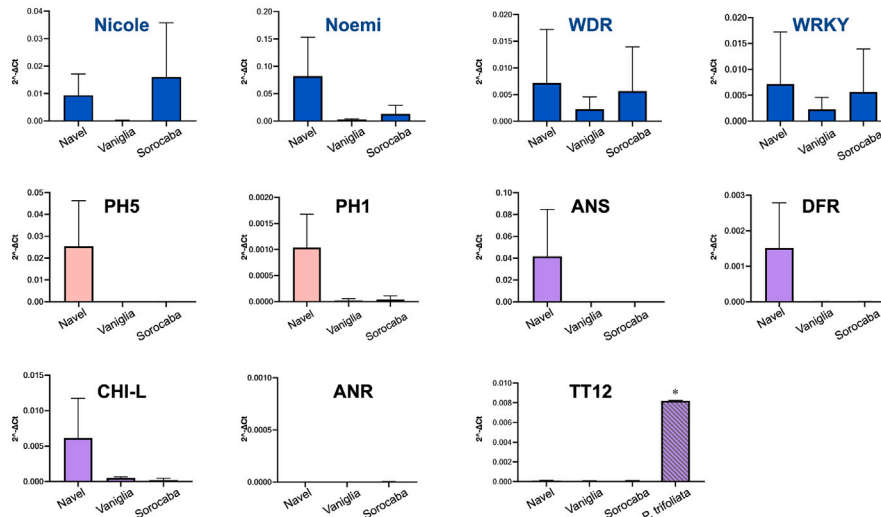
Figure 2. Nicole and Noemi regulate common processes and activate genes required for vacuolar hyperacidification and PA biosynthesis

(A) Alluvial plot illustrating the relationships between genes downregulated in *nicole* ("Lima Verde" and "Lima Sorocaba") and *noemi* ("Vaniglia") mutants compared to wild type ("Navel"). Differentially expressed genes from RNA-seq data were ranked by fold change. Curved lines in color connect identical genes. *PH5* and *PH1* are indicated in light red; PA-related genes in purple. The top eight ranked genes are in common between the two sets of mutants.

(B) Expression analysis of the regulatory genes encoding transcription factors forming the MBW complex and validation of changes observed in RNA-seq data for genes encoding the P-type ATPases PH5 and PH1and genes involved in PA biosynthesis. Note that *ANR* is not expressed in any of the sweet orange varieties analyzed. For *TT12*, the expression in *P. trifoliata* is shown for comparison. Gene expression was determined by RT-qPCR and normalized to the reference gene, actin. Error bars represent the standard deviation of the mean (n = 3 except for *P. trifoliata*, n = 2). Asterisks indicate comparative statistical significance relative to "Navel": * $p < 0.01$.

See also Figure S3 and Table S1.

citrus fruits, we searched natural diversity for varieties with a strong reduction in fruit acidity but, unlike "acidless" pleiotropic *noemi* mutants, still able to produce PAs in seeds. Brazil is one of the world's largest citrus producers and several varieties of sweet orange with very low acidity are relatively popular and collectively referred to as "Lima" (Figure S1F). The most common, simply named "Lima", is a typical "acidless" variety reported as a *noemi* mutant[13] (Figure S1H). We also examined two less common varieties, "Lima Verde" and "Lima Sorocaba" with very low fruit acidity (pH around 6) and citrate content comparable to the "acidless" *noemi* mutant (Figure 1C), but without any impairment in PA production in seeds (Figure S1G).

Sweet orange (*C. sinensis*) is an apomictic hybrid, whose complex mandarin x pummelo genetic constitution is fixed within the population. The absence of sexual recombination and its very narrow genetic base mean that the remarkable phenotypic diversity among accessions results from somatic mutations and that even single polymorphisms in a candidate gene are significant. When analyzed for genes encoding MYB proteins homologous to PhPH4, both "Lima Verde" and "Lima Sorocaba" showed the presence of a defective allele, whose coding sequence was interrupted by a partial LTR retrotransposon of 2,434 bp (Figures 1B and S2A and Data S1). We named this gene *Nicole*. No wild-type allele was detected in the mutants. The absence of a T/G polymorphism downstream of the stop codon (Figure S2B) and the localization of *Nicole* at the 5' end of chromosome 2 (Figure S2C), in an unphased region of apparent mandarin homozygosity,[25] suggest the presence of a terminal deletion affecting the other allele in the mutants. The retrotransposon insertion introduces a premature stop codon that prevents the transcription of the last 188 bp of the *Nicole* coding sequence, without affecting the 5' region encoding the DNA-binding domain of the MYB factor but causing a loss of 62 amino acids which include the putative transcriptional activation domain.

Nicole directly activates the expression of genes required for vacuolar hyperacidification and biosynthesis of proanthocyanidins, which are absent in citrus fruit

To characterize further the *nicole* mutants, we analyzed the accumulation pattern of flavonoids and phenolic acids in fruit juice. Other than a strong reduction in ferulic acid, we could not detect major differences compared to the acidic "Navel" variety (Figure S2D).

We then analyzed the transcriptome of the juice vesicles of fruit of three *nicole* mutants (one accession of "Lima Sorocaba" and two accessions of "Lima Verde" obtained from different sources) and compared these with "Navel" as a wild-type control for acidity, and with the *noemi* "acidless" mutant "Vaniglia".[12] We found that a total of 1,151 genes were downregulated (*p* value <0.05) in all three *nicole* mutants, 636 of which were shared by "Vaniglia", indicating that Nicole and Noemi regulate common processes. Remarkably, the top eight downregulated genes were the same in the two sets of mutants (Figure 2A). Among them were *PH5*, whose expression was completely abolished in both *noemi* and *nicole* mutants and anthocyanin synthase (*ANS*), whose apparently paradoxical expression in fruit of wild-type orange had been reported previously.[26] *ANS* is exclusively involved in the biosynthesis of anthocyanins and PAs (Figure S3A). Generally, anthocyanins are absent in citrus fruit[15] which are also known to lack PAs and their monomers; they are listed as such by USDA, and considered safe for the tannin intolerant.[27,28] Our analyses confirmed the absence of PAs in flesh and juice of sweet orange (Figures S3B–S3F).

Expression analysis by quantitative real-time reverse-transcription PCR (RT-qPCR) (Figure 2B) confirmed that neither of the genes encoding P-type ATPases, *PH5* and *PH1*, were expressed in the *nicole* ("Lima Sorocaba") or *noemi* ("Vaniglia") orange mutants, unlike in wild type "Navel", providing conclusive evidence that fruit hyperacidification is controlled by the Nicole-Noemi complex. *ANS* displayed a similar behavior, and so did *DFR*, another structural gene exclusively involved in anthocyanin and PA biosynthesis. A chalcone isomerase-like gene (*CHI-L*) required for efficient accumulation of PAs in Arabidopsis[29] was downregulated in both *nicole* and *noemi* mutants. Among other PA biosynthetic genes, we detected no expression of *LAR* and *ANR* in fruit of any orange varieties while a gene encoding a PA MATE transporter homologous to TT12[30] was barely expressed in any of the genotypes examined.

Transient transactivation in *N. benthamiana* using a dual luciferase reporter assay demonstrated the ability of the Nicole-Noemi complex to activate the promoters of *PH5*, *PH1*, *ANS*, and *DFR* (Figure 3A). The transcription factors had no activity on the *ANR* promoter. In the same experiments, the truncated version of *Nicole* (*nicole*ˢ), represented by the coding sequence isolated from "Lima Sorocaba" juice, was only marginally active, confirming the effect of the transposable element insertion on the activity of *Nicole*. Iris and Noemi activated the *ANR* promoter of sweet orange (Figure 3A), indicating that the accumulation of PAs in the seed coat of sweet orange, including "Lima Sorocaba"and "Lima Verde" is controlled by the Iris-Noemi complex.

Nicole induces proanthocyanidin accumulation when expressed in tobacco

During the transient assays we also tested the impact of Nicole and Noemi on acidification of leaf tissues, but we found no evidence of a change in pH (Figures S3G and S3H) although we did detect the production of significant amounts of PAs by staining crude extracts of

A

Promoter transactivation in *N. benthamiana*

B

Transient expression in *N. benthamiana*

C
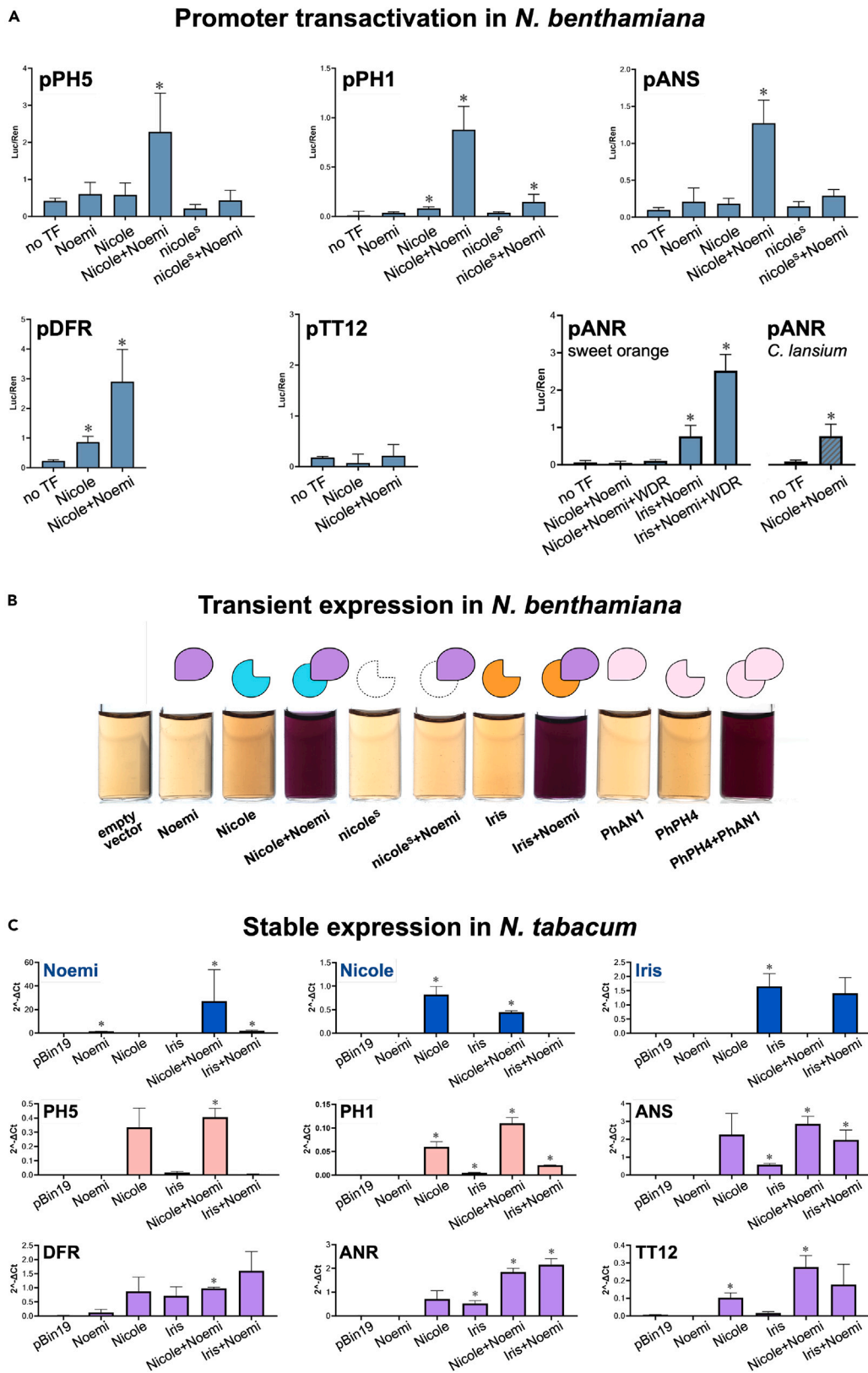
Stable expression in *N. tabacum*

Figure 3. The Nicole-Noemi complex activates the citrus promoters of PA biosynthetic genes and induces PA production when expressed in tobacco

(A) Transactivation assays using the Dual-Luciferase reporter system showing the ability of the Nicole-Noemi complex to activate the citrus promoters of *PH5*, *PH1*, and several PA biosynthetic genes. The mutated protein with a C-terminal deletion, nicole[S], has little or no effect. Note the inability of Nicole-Noemi complex to activate *pANR* from sweet orange (both *pANR[1]* and *pANR[2]* tested) even in the presence of WDR. Error bars represent the standard deviation of the mean (*n* = 5). Asterisks indicate comparative statistical significance relative to "no TF (transcription factor)" control: * *p* < 0.01.

(B) Transient expression of *Nicole* and *Noemi* in *N. benthamiana* results in the accumulation of DMACA-positive metabolites; extracts of infected leaves were mixed mix 0.5 volumes of 0.3% DMACA reagent. A purple pigmentation indicates the production of PAs or their monomers.

(C) Comparative analysis of tobacco plants transformed with the regulatory MYB genes *Nicole* or *Iris* with or without coexpression of the bHLH *Noemi*. Nicole can activate the P-type ATPases *PH5* and *PH1* and all the genes required for PA biosynthesis. Gene expression was determined by RT-qPCR and normalized to the reference gene, actin. Error bars represent the standard deviation of the mean (*n* = 3). Asterisks indicate comparative statistical significance relative to "pBin19" control: * *p* < 0.01.

See also Figure S3, Table S1 and Data S2.

leaf tissue around inoculation sites with DMACA (Figure 3B). In contrast, the defective *Nicole* allele was unable to induce the production of DMACA-positive metabolites after transient expression in *N. benthamiana*. Interestingly, leaves infiltrated with *PhPH4* and *PhAN1*, the known regulators of vacuolar acidity in petunia, produced the same DMACA-positive stain, which was also induced by the combination of Iris and Noemi (Figure 3B).

When stably expressed in *N. tabacum*, *Nicole* induced the production of PAs (Figure S3I) and, correspondingly, the expression of PA biosynthetic genes (Figure 3C). *PH5* and *PH1* were also strongly upregulated. There was little or no activity on the expression of two genes exclusively involved in anthocyanin biosynthesis, 3-GT and PAT (Figure S3A), suggesting that *Nicole* retains its specificity as a PA activator when expressed in tobacco.

Overall, our data in tobacco indicate that the Nicole-Noemi complex directly activates not only *PH1* and *PH5*, key genes required for vacuolar hyperacidification but also genes involved in PA biosynthesis. However, while the entire PA pathway is activated when *Nicole* and *Noemi* are overexpressed in tobacco, only a subset of genes are targets of the complex in sweet orange.

ANR, a gene essential for proanthocyanidin biosynthesis, is not expressed in fruit of sweet orange

In sweet orange, three genes required for PA biosynthesis and accumulation are not expressed in wild-type fruit: *LAR*, *TT12*, and *ANR*. We hypothesized that, during citrus evolution, one of these genes became irresponsive to the Nicole-Noemi complex, favoring additional mutations to arise in other PA genes whose expression became unnecessary in fruit.

LAR is not essential for PA biosynthesis since it is absent in the genome of some species, such as Arabidopsis, which can still produce PAs, formed exclusively on epicatechin units. Furthermore, *LAR* is highly expressed in fruit of citron (*C. medica*) and dependent on the activity of *Noemi* as indicated by the lack of expression in a previously characterized, "acidless" variety[12] (Figures S4A and S4B).

The expression of the transporter *TT12* is negligible in orange. However, we measured high expression in fruit of trifoliate orange (*Poncirus trifoliata*) (Figures 2B, S4C, and S4D), an important species sexually compatible with citrus and now classified as a pure species within the *Citrus* genus.[31] Trifoliate orange is characterized by very acidic fruit.

ANR is the only PA biosynthetic gene that is not expressed in any of the citrus species tested. Unlike *LAR*, *ANR* has been argued to be essential for PA biosynthesis and mutations in this gene result in large reductions in seed coat PAs in Arabidopsis[32] and Medicago.[33] However, it remains unclear whether this is a universal feature in all plants that make PAs.[34]

We observed that mature seeds of trifoliate orange are DMACA negative (Figure S4E), and the *ANR* gene is disrupted in this species, where a region of 2.5 kb spanning the last five exons is replaced by a 2.7 kb insertion corresponding to a repetitive element (Figure S4F). This observation demonstrated that *ANR* is essential for PA biosynthesis in this species. These findings suggested that it was the loss of activity of *ANR* that was the key to the loss of PAs in fruit during citrus evolution. In an ancestral Rutaceae, Nicole may have lost the ability to activate *ANR*, while retaining its ability to activate other PA biosynthetic genes including *PH5* and *PH1*. We speculate that the consequent hyperacidification of the vacuole could not be balanced by the H[+] antiporter activity of TT12, since this transporter is dependent on the production of glycosylated epicatechin PA monomers,[30,35] for which *ANR* is essential[36] (Figure 4).

ANR is expressed in fruit of distant citrus relatives, which also accumulates proanthocyanidins

To test this hypothesis, we searched for species related to citrus, within the family Rutaceae, producing fruits which have retained a functional PA biosynthetic pathway and tested whether these had lower fruit acidity than citrus. The genus *Citrus* belongs to the subfamily Aurantioideae, which is composed of two tribes: Clauseneae and Citreae (to which *Citrus* belongs)[37] (Figure S4G). We obtained fruits from species representative of both tribes, and we tested them for PA production after DMACA staining. As for *Citrus*, the three species within the Citreae tribe, *Atalantia ceylanica*, *Aegle marmelos*, and *Severinia buxifolia*, produced DMACA-negative fruit, but there was strong staining of fruit from the more distantly related members of the Clauseneae tribe: *Glycosmis pentaphylla*, *Clausena anisata*, and *Clausena lansium* (Figure S4H). Fruit of these species had pH in the range 3.7–5.5. We took advantage of the availability of the genome of *Clausena lansium*[38] to isolate the sequence of the *ANR* gene and its promoter. The coding sequences shared 95% nucleotide identity with the two *ANR* alleles of orange but differed considerably in the size of the first intron and in their promoter regions (Data S2). In agreement with the detection of PAs, *ANR* was expressed in fruit of *Clausena lansium* in contrast to orange (Figure S4I) and, unlike orange, the promoter of *ANR* from *Clausena lansium* could be activated by the Nicole-Noemi complex (Figure 3A). We propose a model where *Citrus* may have lost the ability to produce
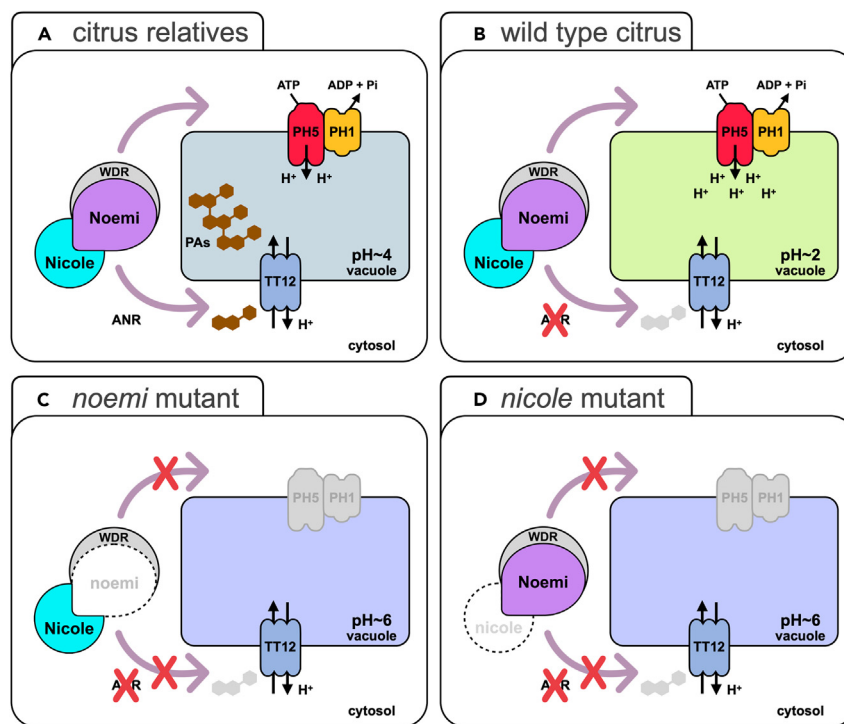
Figure 4. Simplified model of vacuolar hyperacidification in fruit of different citrus accessions and distant relatives

(A) In distant relatives of the *Citrus* genus, the Nicole-Noemi complex activates the expression of PA biosynthetic genes and genes encoding the P-type ATPases PH1 and PH5, resulting in vacuolar hyper acidification required for uptake of PA monomers through the H^+ antiporter TT12.

(B) In citrus, loss of Nicole-Noemi activation of *ANR*, which is essential for PA biosynthesis, and low expression of other biosynthetic genes abolish PA production and generate extreme vacuolar hyper acidification which is not compensated by the import of PA monomers.

(C and D) In citrus mutants *noemi* or *nicole*, the upregulation of the redundant PA biosynthetic genes and the genes encoding P-type ATPases is completely abolished, preventing any possible hyperacidification. In these mutants, strong reduction in vacuolar acidity prevents the movement of citrate into the vacuole down its concentration gradient.

See also Figure S4.

PAs in fruit as a result of a change in the ability of Nicole to activate *ANR*. The inability to produce PAs and the retention of activation of the proton pumps on the tonoplast may have led to hyperacidification through the failure of the MATE proton antiporter to release the proton gradient in the absence of its cargo (Figure 4). This hyperacidification may have been reinforced by the low expression of *DFR* and by the loss of Nicole-Noemi activation of *TT12*, since the equivalent transcription factors in Arabidopsis, AtMYB5 and TT8 can activate *TT12* expression, as well as, at least weakly, activate *ANR*.[20] The ability to produce PAs in fruit has been retained in distantly related species within the same Aurantinoideae subfamily and is associated with lower fruit acidity.

DISCUSSION

The identification and characterization of natural mutants of sweet orange has provided proof that *Nicole* encodes a key transcription factor controlling fruit acidity in citrus. Our study uncovered an extensive overlap between fruit acidity and PA biosynthesis.

In seeds, we have identified a different MYB transcription factor, Iris, controlling production of PAs. When overexpressed in tobacco, both *Nicole* and *Iris* promote PA accumulation through transcriptional activation of all the biosynthetic genes, but only Nicole strongly upregulates the genes encoding the P-type ATPases PH5 and PH1.

In general, the presence of PAs in seeds is almost a universal feature in higher plants and it does not strictly depend on the activity of PH5 and PH1. In fact, plants that do not have either of the corresponding genes in their genomes, for example tomato, are fully capable of producing seed PAs. Nevertheless, the co-regulation of genes required for vacuolar hyperacidification with those involved in PA biosynthesis is not surprising. In Arabidopsis (which does not have a functional PH1), the gene encoding the P-type ATPase equivalent to PH5, *AHA10*, is activated by TT2 during PA biosynthesis, as well as by AtMYB5, encoded by the ortholog of *Nicole*, which has been reported to regulate PA biosynthesis when expressed in the seed coat, through its ability to partially complement *tt2* mutants.[20] This coregulation is plausible since the transport of PA precursors into the vacuole depends on the electrochemical gradient generated by PH5[35,39,40] required to energize the H^+ antiporter TT12. Considerable intersection between these two processes has been documented in other species[41–43] including grape, where two homologs of Nicole activate, with different efficiency, partially overlapping sets of genes.[44]

In the plant kingdom, the presence of PAs in tissues other than seeds is less common and, according to the analysis of the distribution of *PH1* and *PH5* among plant species,[45] appears to be strictly associated with the presence of both ATPases. Fruit of many plants as diverse as apple (dicot) and banana (monocot) are known for their PA content and so the receptacle of strawberry and the leaves of some legumes. All these plant species contains both *PH1* and *PH5*, causing vacuolar hyperacification which may be required not only for the transport of PA monomers across the tonoplast but also for their condensation into polymeric PAs, the final and yet poorly understood step in PA biosynthesis.[34]

Citrus fruit, despite having a functional PA pathway, *PH1*, *PH5*, and expressing *Nicole*, the essential regulator of both sets of genes, do not produce fruit PAs. We associated this feature with the inability to activate *ANR*, the only PA biosynthetic gene for which we could not detect expression in any of the citrus fruit analyzed. The activation of *ANR* by the Nicole-Noemi complex has been retained in fruit of distant citrus relative which, accordingly, accumulate PAs and are also characterized by lower fruit acidity compared to citrus.

We speculate that, in citrus, loss of ANR activity in fruit juice vesicles results in vacuole hyperacidification that is not counteracted by the import of PA precursors. Since plant vacuolar pH is tightly homeostatically regulated, other mechanisms must be in place to ensure that the very steep electrochemical H^+ gradient across the tonoplast is maintained during fruit ripening. Our study, however, indicates that loss of induction of PA biosynthesis may be an integral component of the hyperacidification mechanism in citrus fruit.

Limitations of the study

We propose an evolutionary scenario where the molecular mechanism required for the extreme hyperacidification in citrus fruit may have been co-opted from the proanthocyanin biosynthetic pathway. Although the results presented are consistent with this hypothesis, we do not have direct evidence to support our model (Figure 4B). In particular, we could not establish that lack of ANR activity prevents the efflux of protons induced for PA antiport into the vacuoles of juice sacs, resulting in their hyperacidification. The generation of a knockout *anr* mutant in *N. benthamiana* followed by overexpression of *Nicole* and *Noemi* provided inconclusive results because of a failure of the mutation to inhibit PA accumulation. Further attempts to measure the vacuolar pH in the cells of the inner seed coat of an *anr* mutant of Arabidopsis were unsuccessful because the pH-sensitive fluorescent dye used was unable to permeate this cell layer.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact Eugenio Butelli: eugenio.butelli@jic.ac.uk.

Materials availability

DNA constructs and seeds of transgenic plants generated in this study are available from the lead contact upon request.

Data and code availability

- DNA sequences and RNA-seq reads were deposited into NCBI as indicated in "key resources table".
- All data reported and additional information required to reanalyze the data reported in this paper are available from the lead contact upon request.
- This paper does not report original code.

AUTHOR CONTRIBUTIONS

E.B. and C.M. planned and designed the research; E.B., E.A., E.S., C.C, E.W., L.H., and C.C.H. performed experiments; L.P. C.L., and A.G.-L. provided plant material and information; 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.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  ○ Plant material and growth conditions
- METHOD DETAILS
  ○ Plant material
  ○ pH and citric acid analyses
  ○ Isolation of *Nicole* alleles

- ○ Phylogenetic analysis of citrus MYB transcription factors
- ○ RNA extraction from sweet orange fruit and sequencing
- ○ Real-time quantitative PCR analysis
- ○ Transient expression in *N. Benthamiana* and dual luciferase assay
- ○ pH imaging in *N. Benthamiana*
- ○ Stable transformation in *N. tabacum* and expression analysis
- ○ Localisation and quantification of proantocyanidins
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110923.

REFERENCES

1. Sadka, A., Shlizerman, L., Kamara, I., and Blumwald, E. (2019). Primary Metabolism in Citrus Fruit as Affected by Its Unique Structure. Front. Plant Sci. *10*, 1167. https://doi.org/10.3389/fpls.2019.01167.

2. Burns, J.K., Achor, D.S., and Echeverria, E. (1992). Ultrastructural Studies on the Ontogeny of Grapefruit Juice Vesicles (Citrus-Paradisi Macf Cv Star Ruby). Int. J. Plant Sci. *153*, 14–25. https://doi.org/10.1086/297002.

3. 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. https://doi.org/10.1038/nature25447.

4. Wang, L., He, F., Huang, Y., He, J., Yang, S., Zeng, J., Deng, C., Jiang, X., Fang, Y., Wen, S., et al. (2018). Genome of Wild Mandarin and Domestication History of Mandarin. Mol. Plant *11*, 1024–1037. https://doi.org/10.1016/j.molp.2018.06.001.

5. Martinoia, E., Maeshima, M., and Neuhaus, H.E. (2007). Vacuolar transporters and their essential role in plant metabolism. J. Exp. Bot. *58*, 83–102. https://doi.org/10.1093/jxb/erl183.

6. Etienne, A., Génard, M., Lobit, P., Mbeguié-A-Mbéguié, D., and Bugaud, C. (2013). What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. J. Exp. Bot. *64*, 1451–1469. https://doi.org/10.1093/jxb/ert035.

7. de Vlaming, P., Schram, A.W., and Wiering, H. (1983). Genes affecting flower colour and pH of flower limb homogenates in Petunia hybrida. Theor. Appl. Genet. *66*, 271–278. https://doi.org/10.1007/BF00251158.

8. 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. https://doi.org/10.1105/tpc.105.034041.

9. Verweij, W., Spelt, C., Di Sansebastiano, G.P., Vermeer, J., Reale, L., Ferranti, F., Koes, R., and Quattrocchio, F. (2008). An H+ P-ATPase on the tonoplast determines vacuolar pH and flower colour. Nat. Cell Biol. *10*, 1456–1462. https://doi.org/10.1038/ncb1805.

10. Shi, C.Y., Song, R.Q., Hu, X.M., Liu, X., Jin, L.F., and Liu, Y.Z. (2015). Citrus PH5-like H(+)-ATPase genes: identification and transcript analysis to investigate their possible relationship with citrate accumulation in fruits. Front. Plant Sci. *6*, 135. https://doi.org/10.3389/fpls.2015.00135.

11. Shi, C.Y., Hussain, S.B., Yang, H., Bai, Y.X., Khan, M.A., and Liu, Y.Z. (2019). CsPH8, a P-type proton pump gene, plays a key role in the diversity of citric acid accumulation in citrus fruits. Plant Sci. *289*, 110288. https://doi.org/10.1016/j.plantsci.2019.110288.

12. Butelli, E., Licciardello, C., Ramadugu, C., Durand-Hulak, M., Celant, A., Reforgiato Recupero, G., Froelicher, Y., and Martin, C. (2019). Noemi Controls Production of Flavonoid Pigments and Fruit Acidity and Illustrates the Domestication Routes of Modern Citrus Varieties. Curr. Biol. *29*, 158–164.e2. https://doi.org/10.1016/j.cub.2018.11.040.

13. Strazzer, P., Spelt, C.E., Li, S., Bliek, M., Federici, C.T., Roose, M.L., Koes, R., and Quattrocchio, F.M. (2019). Hyperacidification of Citrus fruits by a vacuolar proton-pumping P-ATPase complex. Nat. Commun. *10*, 744. https://doi.org/10.1038/s41467-019-08516-3.

14. Gonzalez, A., Mendenhall, J., Huo, Y., and Lloyd, A. (2009). TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. Dev. Biol. *325*, 412–421. https://doi.org/10.1016/j.ydbio.2008.10.005.

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. https://doi.org/10.1105/tpc.111.095232.

16. 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. https://doi.org/10.1104/pp.16.01701.

17. Dutt, M., Mahmoud, L.M., Weber, K.C., Satpute, A., Stanton, D., Qiu, W., Soares, J.M., Reuss, L., Wang, Y., Grosser, J.W., and Killiny, N. (2023). Anthocitrus: evaluation of anthocyanin accumulating "Mexican" lime fruits produced by overexpressing the Ruby transcription factor gene from Citrus sinensis "Moro". Plant Cell Tiss Org *155*, 283–296. https://doi.org/10.1007/s11240-023-02581-9.

18. Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell *13*, 2099–2114. https://doi.org/10.1105/tpc.13.9.2099.

19. Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. Plant J. *39*, 366–380. https://doi.org/10.1111/j.1365-313X.2004.02138.x.

20. Xu, W., Grain, D., Bobet, S., Le Gourrierec, J., Thévenin, J., Kelemen, Z., Lepiniec, L., and Dubos, C. (2014). Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. New Phytol. *202*, 132–144. https://doi.org/10.1111/nph.12620.

21. Bogs, J., Jaffé, F.W., Takos, A.M., Walker, A.R., and Robinson, S.P. (2007). The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. Plant Physiol. *143*, 1347–1361. https://doi.org/10.1104/pp.106.093203.

22. Zhang, Y., Ye, J., Liu, C., Xu, Q., Long, L., and Deng, X. (2020). Citrus PH4-Noemi regulatory complex is involved in proanthocyanidin biosynthesis via a positive feedback loop. J. Exp. Bot. *71*, 1306–1321. https://doi.org/10.1093/jxb/erz506.

23. Li, S.J., Liu, X.J., Xie, X.L., Sun, C.D., Grierson, D., Yin, X.R., and Chen, K.S. (2015). CrMYB73, a PH-like gene, contributes to citric acid accumulation in citrus fruit. Sci. Hortic. *197*, 212–217. https://doi.org/10.1016/j.scienta.2015.09.037.

24. Huang, Y., He, J., Xu, Y., Zheng, W., Wang, S., Chen, P., Zeng, B., Yang, S., Jiang, X., Liu, Z., et al. (2023). Pangenome analysis provides insight into the evolution of the orange subfamily and a key gene for citric acid accumulation in citrus fruits. Nat. Genet. *55*, 1964–1975. https://doi.org/10.1038/s41588-023-01516-6.

25. Wu, B., Yu, Q., Deng, Z., Duan, Y., Luo, F., and Gmitter, F., Jr. (2023). A chromosome-level phased genome enabling allele-level studies

in sweet orange: a case study on citrus Huanglongbing tolerance. Hortic. Res. *10*, uhac247. https://doi.org/10.1093/hr/uhac247.

26. Huang, D., Zhao, Y., Cao, M., Qiao, L., and Zheng, Z.L. (2016). Integrated Systems Biology Analysis of Transcriptomes Reveals Candidate Genes for Acidity Control in Developing Fruits of Sweet Orange (Citrus sinensis L. Osbeck). Front. Plant Sci. *7*, 486. https://doi.org/10.3389/fpls.2016.00486.

27. Gu, L., Kelm, M.A., Hammerstone, J.F., Beecher, G., Holden, J., Haytowitz, D., and Prior, R.L. (2003). Screening of foods containing proanthocyanidins and their structural characterization using LC-MS/MS and thiolytic degradation. J. Agric. Food Chem. *51*, 7513–7521. https://doi.org/10.1021/jf034815d.

28. Haytowitz, D., Wu, X., and Bhagwat, S. (2018). USDA Database for the Proanthocyanidin Content of Selected Foods 2018.

29. Waki, T., Mameda, R., Nakano, T., Yamada, S., Terashita, M., Ito, K., Tenma, N., Li, Y., Fujino, N., Uno, K., et al. (2020). A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. Nat. Commun. *11*, 870. https://doi.org/10.1038/s41467-020-14558-9.

30. Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J.M., Debeaujon, I., and Klein, M. (2007). The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. Plant Cell *19*, 2023–2038. https://doi.org/10.1105/tpc.106.046029.

31. Ollitrault, P., Curk, F., and Krueger, R. (2020). Chapter 4 - Citrus taxonomy. In The Genus Citrus, M. Talon, M. Caruso, and F.G. Gmitter, eds. (Woodhead Publishing), pp. 57–81. https://doi.org/10.1016/B978-0-12-812163-4.00004-8.

32. Albert, S., Delseny, M., and Devic, M. (1997). BANYULS, a novel negative regulator of flavonoid biosynthesis in the Arabidopsis seed coat. Plant J. *11*, 289–299. https://doi.org/10.1046/j.1365-313x.1997.11020289.x.

33. Liu, C., Wang, X., Shulaev, V., and Dixon, R.A. (2016). A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. Nat. Plants *2*, 16182. https://doi.org/10.1038/nplants.2016.182.

34. Yu, K., Song, Y., Lin, J., and Dixon, R.A. (2023). The complexities of proanthocyanidin biosynthesis and its regulation in plants. Plant Commun. *4*, 100498. https://doi.org/10.1016/j.xplc.2022.100498.

35. Zhao, J., and Dixon, R.A. (2009). MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in Medicago truncatula and Arabidopsis. Plant Cell *21*, 2323–2340. https://doi.org/10.1105/tpc.109.067819.

36. Jun, J.H., Lu, N., Docampo-Palacios, M., Wang, X., and Dixon, R.A. (2021). Dual activity of anthocyanidin reductase supports the dominant plant proanthocyanidin extension unit pathway. Sci. Adv. *7*, eabg4682. https://doi.org/10.1126/sciadv.abg4682.

37. Penjor, T., Yamamoto, M., Uehara, M., Ide, M., Matsumoto, N., Matsumoto, R., and Nagano, Y. (2013). Phylogenetic relationships of citrus and its relatives based on matK gene sequences. PLoS One *8*, e62574. https://doi.org/10.1371/journal.pone.0062574.

38. Fan, Y., Sahu, S.K., Yang, T., Mu, W., Wei, J., Cheng, L., Yang, J., Liu, J., Zhao, Y., Lisby, M., and Liu, H. (2021). The Clausena lansium (Wampee) genome reveal new insights into the carbazole alkaloids biosynthesis pathway. Genomics *113*, 3696–3704. https://doi.org/10.1016/j.ygeno.2021.09.007.

39. 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. https://doi.org/10.1007/s10142-011-0226-3.

40. Appelhagen, I., Nordholt, N., Seidel, T., Spelt, K., Koes, R., Quattrochio, F., Sagasser, M., and Weisshaar, B. (2015). TRANSPARENT TESTA 13 is a tonoplast P3A -ATPase required for vacuolar deposition of proanthocyanidins in Arabidopsis thaliana seeds. Plant J. *82*, 840–849. https://doi.org/10.1111/tpj.12854.

41. Deluc, L., Bogs J Fau - Walker, A.R., Walker Ar Fau - Ferrier, T., Ferrier T Fau - Decendit, A., Decendit A Fau - Merillon, J.-M., Merillon Jm Fau - Robinson, S.P., Robinson Sp Fau - Barrieu, F., and Barrieu, F. (2008). The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiol. *147*, 2041–2053.

42. Liu, C., Jun, J.H., and Dixon, R.A. (2014). MYB5 and MYB14 Play Pivotal Roles in Seed Coat Polymer Biosynthesis in Medicago truncatula. Plant Physiol. *165*, 1424–1439. https://doi.org/10.1104/pp.114.241877.

43. Jiang, L., Yue, M., Liu, Y., Zhang, N., Lin, Y., Zhang, Y., Wang, Y., Li, M., Luo, Y., Zhang, Y., et al. (2023). A novel R2R3-MYB transcription factor FaMYB5 positively regulates anthocyanin and proanthocyanidin biosynthesis in cultivated strawberries (Fragaria × ananassa). Plant Biotechnol. J. *21*, 1140–1158.

44. Amato, A., Cavallini, E., Walker, A.R., Pezzotti, M., Bliek, M., Quattrocchio, F., Koes, R., Ruperti, B., Bertini, E., Zenoni, S.A.-O., and Tonielli, G.B. (2019). The MYB5-driven MBW complex recruits a WRKY factor to enhance the expression of targets involved in vacuolar hyper-acidification and trafficking in grapevine. Plant J. *99*, 1220–1241.

45. Li, Y., Provenzano, S., Bliek, M., Spelt, C., Appelhagen, I., Machado de Faria, L., Verweij, W., Schubert, A., Sagasser, M., Seidel, T., et al. (2016). Evolution of tonoplast P-ATPase transporters involved in vacuolar acidification. New Phytol. *211*, 1092–1107. https://doi.org/10.1111/nph.14008.

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

47. Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods *1*, 13.

48. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.

49. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

50. Alexa, A., and Rahnenfuhrer, J. (2023). topGO: Enrichment Analysis for Gene Ontology. R package version 2.54.0.

51. Powell, D., Milton, M., Perry, A., and Santos, K. (2019). Drpowell/Degust 4.1. 1 (Zenodo).

52. Lin, Q., Wang, C., Dong, W., Jiang, Q., Wang, D., Li, S., Chen, M., Liu, C., Sun, C., and Chen, K. (2015). Transcriptome and metabolome analyses of sugar and organic acid metabolism in Ponkan (Citrus reticulata) fruit during fruit maturation. Gene *554*, 64–74. https://doi.org/10.1016/j.gene.2014.10.025.

53. Catalano, C., Ciacciulli, A., Salonia, F., Russo, M.P., Caruso, P., Caruso, M., Russo, G., Distefano, G., and Licciardello, C. (2020). Target-Genes Reveal Species and Genotypic Specificity of Anthocyanin Pigmentation in Citrus and Related Genera. Genes *11*, 807. https://doi.org/10.3390/genes11070807.

54. Gambino, G., Perrone, I., and Gribaudo, I. (2008). A Rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. Phytochem. Anal. *19*, 520–525. https://doi.org/10.1002/pca.1078.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
A. tumefaciens LBA4404	Lab stock	N/A
A. tumefaciens GV3101	Lab stock	N/A
Chemicals, peptides, and recombinant proteins		
4-Dimethylaminocinnamaldehyde (DMACA)	Sigma-Aldrich	Cat # D4506
(+)-Catechin	Sigma-Aldrich	Cat # 43412-10MG
(−)-epicatechin	Sigma-Aldrich	Cat # 68097-10MG
6-carboxyfluorescein diacetate (6-CFDA)	Sigma-Aldrich	Cat # C5041
Critical commercial assays		
DNeasy Plant Mini Kit	QIAGEN	Cat # 69106
RNeasy Plant Mini Kit	QIAGEN	Cat # 74904
Phusion High-Fidelity DNA Polymerase	Thermo Fisher	Cat # F530L
SuperScript IV Reverse Transcriptase	Thermo Fisher	Cat # 18090050
SYBR Green JumpStart Taq ReadyMix Kit	Sigma-Aldrich	Cat # S4438-500RXN
Dual-Glo Luciferase Assay System	Promega	Cat # E2920
Custom DNA sequencing	Eurofins	https://www.eurofinsgenomics.eu/
Deposited data		
Nucleotide sequence of wild type *Nicole* in sweet orange and downstream T/G polymorphism	This study	GenBank: OR_766345
Nucleotide sequence of mutated *nicole* in sweet orange "Lima Verde" and "Lima Sorocaba"	This study	GenBank: OR_766346
RNA-Seq data deposited into NCBI sequence read archive (SRA)	This study	NIH BioProject: PRJNA1071547
Experimental models: Organisms/strains		
Accession of sweet orange (*C. sinensis*)	Listed in 'Plant Material'	N/A
Oligonucleotides		
See Table S2 for primers	This study	N/A
Recombinant DNA		
pDONR 207	Thermo Fisher	N/A
pJAM1502	This study	N/A
pEAQ-HT-DEST1	Sainsbury et al.[46]	N/A
pGreen II 0800-LUC	Hellens et al.[47]	N/A
Software and algorithms		
PhyML	Guindon et al.[48]	http://www.atgc-montpellier.fr/phyml/binaries.php
RAxML	Stamatakis et al.[49]	https://cme.h-its.org/exelixis/web/software/raxml/
topGO	Alexa et al.[50]	https://bioconductor.org/packages/release/bioc/html/topGO.html
DEGUST	Powel et al.[51]	http://degust.erc.monash.edu/

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Citrus Genome Assembly and Annotation	*Citrus sinensis* v1.1	https://phytozome-next.jgi.doe.gov/info/Csinensis_v1_1
Citrus Genome Assembly and Annotation	CPBD: Citrus Pan-genome	http://citrus.hzau.edu.cn/
Citrus Genome Assembly and Annotation	*Citrus sinensis* "Valencia" (set A)[25]	NIH BioProject: PRJNA736174
Citrus Genome Assembly and Annotation	*Citrus sinensis* "Valencia" (set B)[25]	NIH BioProject: PRJNA736176

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plant material and growth conditions

Accessions of sweet orange and citrus relatives were obtained from different sources as indicated in 'method details'. Transgenic tobacco plants were obtained in the Samsun NN line of *Nicotiana tabacum* and kept in greenhouse conditions. Greenhouses were kept to 20°C in the day and 16°C at night. Supplemental lighting was used in winter months, maintaining a 16-h day.

METHOD DETAILS

Plant material

"Navel" and "Valencia" are acidic sweet orange varieties, wild type for acidity used as comparative controls. For RNA-seq analysis, ripe fruit of one "Navel" accession were obtained from CREA (Acireale, Italy). The other biological replicates of "Navel" oranges were obtained from a local supermarket in Norwich (Tesco). "Valencia" fruits were obtained from a local supermarket in Norwich (Waitrose). "Vaniglia" fruits for all the biological replicates were obtained from CREA. "Lima Sorocaba" was obtained from a local market in São Paulo region, Brazil. "Lima Verde" fruits were obtained from Santa Cruz do Rio Pardo (R1) and Mogi Guaçu (R2). Fruits of *Severinia buxifolia* were obtained from CREA. The remaining distant citrus relatives were obtained from Instituto Valenciano de Investigaciones Agrarias (IVIA, Valencia, Spain). *P. trifoliata* plant material was obtained from a tree in Norwich, grown from seeds collected at CREA.

pH and citric acid analyses

Fruit juice was obtained using a conventional citrus squeezer from fruits of sweet orange and *P. trifoliata* and with a plastic pestle from fruit of distant citrus relatives. After centrifugation, pH of juice was measured using a standard combination Ag/AgCl pH electrode in at least three fruits per variety. Citric acid in different sweet orange varieties was quantified by gas chromatography-mass spectrometry (GCMS) as described by Lin et al.[52] Three biological replicates, obtained from individual fruits, were used except for "Navel" where only two samples were available.

Isolation of *Nicole* alleles

Full-length *Nicole* alleles were isolated by PCR using primers PH4-FA and PH4-RZ (Table S2) designed close to start and stop codons of the gene. DNA was isolated from leaves or fruit of different sweet orange varieties using the DNeasy Plant Mini Kit (Qiagen). PCR fragments were used directly for sequencing. Appropriate primers were used to obtain the complete sequences and confirm the presence of the retrotransposon. All the sequences obtained were compared to those available in different citrus genome assemblies listed in the key resources table.

Phylogenetic analysis of citrus MYB transcription factors

Genome-wide analyses of R2R3-MYB proteins in the sweet orange v1.1 Phytozome reference genome were performed by protein sequence analyses, multiple sequence alignments, and subsequent phylogenetic tree construction. Other proteins from different species were also included: 125 R2R3-MYBs from Arabidopsis and characterised R2R3-MYBs from *V. vinifera* (n = 5), *P. hybrida* (n = 2), *L. chinensis* (n = 2), *M. truncatula* (n = 3), *M. domestica* (n = 4), *P. persica* (n = 3), *G. max* (n = 1) and *F. ananassa* (n = 1). Initially, candidate genes encoding MYB TFs were identified by reference genome amino acid sequence analysis of PFAM domains PF00249 (Myb-like DNA binding domain). Candidate sweet orange MYB-like TFs containing two predicted MYB domains were selected and designated as R2R3-CsMYBs (Table S1). The maximum-likelihood (ML) substitution model that best described each alignment was determined using PhyML.[48] Phylogenetic trees were constructed using the best substitution model with RAxML[49] (1,000 bootstrap replicates).

RNA extraction from sweet orange fruit and sequencing

RNA from sweet orange fruits of different varieties was extracted as previously reported[53] and further purified using an RNeasy Plant Mini Kit (Qiagen). RNA samples were sent to Novogene for low input PE150 Illumina sequencing. Novogene conducted the initial analyses, providing FPKM values for all genes in all samples and significance values when compared to expression in wild type "Navel" fruit juice. R (v4.0.5) was used to generate lists of differentially expressed genes in all the mutant varieties and GO enrichment analyses using R package topGO.[50] For

comparative differential expression analyses between wild type and 'acidless' *noemi* mutant or the combined three *nicole* mutants, read counts were analyzed using the interactive web-tool DEGUST using the Voom function from limma package methodo methodology.[51]

Real-time quantitative PCR analysis

Quantitative real-time reverse-transcription PCR (RT-qPCR) was used to validate RNA-seq transcriptomic data and to analyze gene expression in transgenic tobacco. For sweet orange, total RNA was extracted as described above, treated with DNAse I (Roche) and retrotranscribed using Superscript IV reverse transcriptase (Invitrogen). RT-qPCR primers were designed to span exon-exon junctions or were separated by large introns to avoid amplification of residual DNA contamination and tested for primer efficiency. Reactions were conducted using the SYBR Green JumpStart Taq ReadyMix Kit (Sigma) with the X96 Touch Real-Time PCR Detection System (Biorad). Three biological and three technical replicates were used for each variety of sweet orange. For conventional RT-PCR, first-strand cDNA was obtained as described above and amplified with Phusion High-Fidelity DNA Polymerase (Thermo Fisher) and primers indicated in Table S2. DNA bands were sequenced to confirm correct exon-exon junctions.

Transient expression in *N. Benthamiana* and dual luciferase assay

The effect of different transcription factors on PA production and promoter activation was assessed by transient expression via agroinfiltration of leaves of *N. benthamiana*. Effector plasmids encoding transcription factors from sweet orange were obtained by PCR amplification using cDNA from fruit of sweet orange or petals of petunia (V26) and primers listed in Table S2. The coding sequences were first cloned into the pDONR207 entry clone (Thermo Fisher) and subsequently transferred to the destination vector pEAQ-HT-DEST1[46] designed for Gateway cloning technology (Invitrogen). The reporter constructs were obtained by PCR amplification of different promoters followed by restriction enzyme cloning into the pGreen II 0800-LUC vector[47] to control the expression of the firefly-derived luciferase reporter gene. This vector also contained a *Renilla* luciferase gene under the control of a *CaMV 35S* promoter as an internal control to normalise the values of the experimental reporter gene. Effector and reporter plasmids were transformed into *A. tumefaciens* strain GV3101. Reporter plasmids were co-transformed with the helper plasmid pSoup. Liquid cultures were grown overnight with appropriate antibiotic selection and harvested by centrifugation. Cells were washed and resuspended in 10 mM $MgCl_2$, 10 mM MES pH 5.6, 200 μM acetosyringone to A_{600} = 0.1 for effector and A_{600} = 0.2 for reporter constructs. *A. tumefaciens* suspensions were infiltrated into the abaxial surface of expanded leaves of *N. benthamiana*. For each combination, five injected areas were treated as biological replicates. Agroinfiltrated leaves were harvested after 3 days, and luciferase activity was measured immediately using the Dual-Glo Luciferase Assay System kit (Promega). Leaf discs of 4 mm in diameter were collected in 1.5 mL white transparent tubes containing 100 μL of PBS. A volume of 75 μL of luciferase assay reagent was added and firefly luminescence was measured on a Glomax 20/20 single tube luminometer (Promega) after 10 min. *Renilla* luminescence was measured on the same instrument 10 min after the addition to the same sample of 75 μL of Stop & Glo reagent. Results were expressed as the ratio of firefly to *Renilla* luciferase activity (Luc/Ren).

pH imaging in *N. Benthamiana*

The method employed a non-fluorescent dye, 6-carboxyfluorescein diacetate (6-CFDA, Sigma-Aldrich) which is cleaved by non-specific vacuolar esterases releasing the fluorescent dye 6-carboxyfluorescein (6-CF), which has a pH-dependent ratio of emission intensities at different excitation wavelengths. Two days after agroinfiltration, leaf discs were infiltrated with pH staining solution (20 μM 6-CFDA in MS liquid medium). Leaf discs were collected after 24 h and incubated with 13.3 μM 6-carboxyfluorescein di acetate (6-CFDA), 0.3% Triton X-100 in MS liquid medium. Images were captured using a Leica SP5 Confocal Stereo Microscope equipped with an HC PL APO CS2 20x/0.75 air objective (Leica Microsystems GmbH). The absorbance emission derived from 458 nm to 488 nm excitation lasers was measured and the 488/458 nm emission ratio was calculated and compared with a calibration curve obtained with 6-CF dye in buffers with different known pH.

Stable transformation in *N. tabacum* and expression analysis

The coding sequences of different regulatory genes were obtained as described above and cloned into pJAM 1502, a pBin19-derived binary vector equipped with a double *CaMV 35S* promoter and a *CaMV* terminator with *attR* recombination sites in between, using Gateway cloning technology. The resulting plasmids were transferred to A. tumefaciens strain LBA4404 and used to transform N. tabacum cv Samsun NN. Plants transformed with the empty pJAM 1502 were used as a control. For gene expression analysis, total RNA was extracted using an optimised cetyltrimethylammonium bromide (CTAB)-based protocol[54] to minimise the binding of PAs to RNA. Sepals were used for the lines expressing individual regulatory genes and leaves for co-expressing lines, since *35S::Nicole - 35S::Noemi* plants did not produce any flowers. RT-qPCR was conducted as described above. Three biological and three technical replicates were used for each transgenic plant. For the biological replicates, at least two independent transformants or, for the double overexpressors, two independent crosses between different transformants were used for each genotype.

Localisation and quantification of proanthocyanidins

The presence of proanthocyanidins in seeds, fruit, leaves, and flowers was determined by staining with 0.3% (w/v) DMACA (4-Dimethylaminocinnamaldehyde, Sigma) in methanol and 6 M HCl (1:1, v/v) followed by several washing steps with 70% ethanol. For *N. tabacum* leaves and flowers, samples were first destained overnight in a solution of ethanol-glacial acetic acid (3:1, v/v). For transient expression in

N. benthamiana, freeze-dried leaf samples were extracted with 80% methanol overnight and 0.5 vol. of 0.3% DMACA reagent was added. In presence of PAs or their monomers, the solution initially quickly turned blue-green, then purple. If the compounds were not present, the solution remained yellow. For HPLC analysis, PAs were extracted from fruit of different sweet orange varieties and, as controls, from grape skin and transgenic *N. tabacum* plants generated in this study. Samples were extracted twice with 10 vol. of an extraction solution, containing 70% (v/v) acetone and 0.5% (v/v) acetic acid, and sonication at room temperature for 30 min. Extractions were centrifugated at 8,000 rpm for 10 min and supernatants containing soluble PAs were pooled. Soluble PA samples were washed three times with chloroform, and a further three times with hexane, before freeze-drying overnight. PA powder was resuspended in 1 mL extraction solution per 1 g of fresh starting material. The extracts were diluted 5-fold in H_2O and run on a Shimadzu Nexera UHPLC with Prominence diode array detector (UV/vis absorbance) and a 2020 single quad mass spectrometer. Separation was on a 50 × 2.1 mm, 2.6 μm particle size Kinetex EVO C18 column (Phenomenex) using a gradient of acetonitrile versus 0.1% (w/v) formic acid in H_2O, run at 0.6 mL min^{-1} at 40°C. Detection was by UV absorbance collecting full spectra from 200 to 600 nm at 12.5 Hz with a time constant of 0.08 s, and by positive mode electrospray MS. The mass spec collected full spectra from *m/z* 100–900 in 0.1 s and monitored *m/z* 291(+) by single-ion-monitoring for 50 msec. Spray chamber conditions were 250°C desorbation line, 200°C heat block, 1.5 L min^{-1} nebulizer gas, and 15 L min^{-1} drying gas. The injection volume was 5 μL. Flavan-3-ol monomers were quantified at 279 nm by calibration against (+)-catechin and (−)-epicatechin standard curves. Following acid-catalysis of soluble PA extracts, flavan-3-ol terminal subunits from PA polymers were released in addition to extension subunits. Terminal subunit concentration was determined by the subtraction of free flavan-3-ol monomer concentration in uncleaved PA samples from the total flavan-3-ol monomer concentration in the acid-catalysed samples.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bar plots were generated using GraphPad Prism 10 and Excel version 2308. Error bars represent standard deviation of the mean (SD). Asterisks indicate comparative statistical significance relative to "Navel" calculated by unpaired Student's t test.

The number of the replicates assayed is described in the STAR Methods and in the figure legends.