UNIVERSITAT POLITÈCNICA DE VALÈNCIA

MUTATION RATE AND GENE EXPRESSION IN GENETIC ADMIXTURES OF DOMESTICATED CITRUS

PhD Thesis

Estela Pérez Román

Supervised by Dr. Manuel Talón Cubillo

València, November 2022

PhD Thesis

Mutation Rate and Gene Expression in Genetic Admixtures of Domesticated Citrus

For the degree of

Doctor in Biotechnology

Estela Pérez Román

Supervised by Dr. Manuel Talón Cubillo

This doctoral thesis was carried out at the Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias (IVIA)

València, November 2022

PhD Thesis

Mutation Rate and Gene Expression in Genetic Admixtures of Domesticated Citrus

For the degree of

Doctor in Biotechnology

Estela Pérez Román

Supervised by Dr. Manuel Talón Cubillo

This doctoral thesis was funded by Ministerio de Ciencia e Innovación (Agencia Estatal de Investigación – INIA) through grant #RTA2014-00071-C06-01, #CPD2016-0067 (FPI Program)

València, November 2022

Dr. Manuel Talón Cubillo, Director del Centro de Genómica del Instituto Valenciano de Investigaciones Agrarias (IVIA), certifica que,

Dña. Estela Pérez Román, Licenciada en Biotecnología por la Universitat Politècnica de València (UPV), ha realizado bajo su dirección y supervisión el trabajo de investigación y la memoria de la tesis titulada *Mutation Rate and Gene Expression in Genetic Admixtures of Domesticated Citrus*, que presenta para optar al grado de Doctor en Biotecnología por la Universitat Politècnica de València.

Y para que conste a los efectos oportunos, firma el presente certificado en València, octubre de 2022.

MANUEL|TALON|CUBILLO

Firmado digitalmente por MANUEL|TALON|CUBILLO
Fecha: 2022.10.24 14:26:53 +02'00'

Fdo. Dr. Manuel Talón Cubillo

Table of Content

Abstract

Citrus domestication is a largely unknown process. Research carried out to date indicates that current commercial citrus varieties are, in general, the product of ancestral intraspecific and interspecific introgressions. Intraspecific introgressions took place between the populations of two antique subspecies of mandarins and contributed to the apomictic behavior that most current mandarins and other related citrus still retain today. Once these genetic admixtures were established, interspecific pummelo introgression brought new traits to the admixtured mandarin genome. During domestication, genetic variability of these citrus occurred mainly through spontaneous mutations and with the practice of grafting, seedless genotypes expanded rapidly, emphasizing the role of mutations in the selection of citrus such as oranges and modern mandarins.

In this work, we provide evidence that somatic mutations in citrus propagate following an iterative pattern determined by the sympodial branching model and consequently are grouped in sectors along the tree, where some of them remain fixed. In this scenario, a tree can be considered a mosaic genetically composed of different genomes, in which younger branches accumulate greater number of mutations. On average, our 36-yr-old experimental Clementine tree shows a mutation rate of 4.4×10^{-10} bp^{-1} yr^{-1}, may carry a total of 1,500 to 5,000 variants and produces 1 somatic mutation per axillary meristem. This relatively high number of mutations is in line with the large number of varieties derived from spontaneous mutations that are commercialized in citrus.

To identify key traits elicited by citrus domestication, we analyzed transcriptomes from developing fruit pulp of wild inedible Ichang papeda, Sun Chu Sha Kat sour mandarin and three palatable segregants derived from a cross between commercial mandarins. Based on these results, we propose that during the transition from inedible papedas to sour mandarins, domestication involved a first phase of major changes in the expression of genes regulating central pathways of primary and secondary metabolism. This stage was mainly characterized by both the stimulation of growth processes and the reduction of distasteful chemicals defenses. It is also suggested that in a second phase, edible attributes of mandarins, especially acidity, were progressively improved through specific modifications. Other relevant changes included upregulation of genes involved in the synthesis of key substances of pleasant aroma and flavor, and the modulation of

sugar transporters. Thus, the transition from inedible papeda to sour mandarin was essentially defined by a drastic reprogramming of gene expression of fundamental metabolic pathways, while modern mandarins evolved later through progressive refinement of palatability properties. Taken together, these observations suggest that the rates of mutations occurring in citrus somatic tissues are consistent with the idea that they have played an important role during the later steps of citrus domestication.

Resumen

La domesticación de los cítricos es un proceso en gran parte desconocido. Las investigaciones llevadas a cabo hasta la fecha indican que las variedades de cítricos que se comercializan en la actualidad son, en general, el producto de introgresiones ancestrales intraespecíficas e interespecíficas. Las introgresiones de tipo intraespecífico tuvieron lugar entre las poblaciones de dos subespecies antiguas de mandarinas y aportaron el comportamiento apomíctico que la mayoría de las mandarinas actuales y otros cítricos relacionados conservan todavía hoy. Una vez dichas mezclas genéticas o *admixtures* quedaron establecidas, otras introgresiones interespecíficas de *pummelo* establecieron nuevas características en el genoma de mandarina. Durante el proceso de domesticación, la variabilidad genética de estos cítricos se adquirió fundamentalmente a través de mutación espontánea, y mediante la práctica del injerto los genotipos carentes de semillas se extendieron rápidamente, acentuándose el papel de las mutaciones en la selección de cítricos como las naranjas y las mandarinas modernas.

En este trabajo, mostramos que las mutaciones somáticas en cítricos se propagan siguiendo un patrón iterativo determinado por un modelo de ramificación simpodial y, en consecuencia, se agrupan en sectores a lo largo del árbol, donde algunas mutaciones quedan permanentemente fijadas. En este escenario, un árbol puede considerarse un mosaico genéticamente compuesto de diferentes genomas, en el que las ramas más jóvenes acumulan un número mayor de mutaciones. En promedio, nuestro árbol experimental de Clementina de 36 años muestra una tasa de mutación de 4.4×10^{-10} bp^{-1} yr^{-1}, pudiendo contener en total entre 1500 y 5000 variantes y producir 1 mutación somática en cada meristemo axilar. Este número relativamente alto de mutaciones está en línea con el gran número de variedades derivadas de mutaciones espontáneas que se comercializan en cítricos.

Para identificar caracteres esenciales adquiridos durante la domesticación, se han analizado los transcriptomas de la pulpa de frutos en desarrollo de la variedad silvestre e incomestible papeda de Ichang, la mandarina ácida Sun Chu Sha Kat y tres segregantes comestibles derivados de un cruce entre mandarinas comerciales. Basándonos en los resultados obtenidos, se propone que durante la transición de las papedas incomestibles a las mandarinas ácidas, la domesticación se caracterizó por una primera fase de cambios radicales en la expresión de los genes que regulan rutas fundamentales tanto

del metabolismo primario como del secundario. Esta fase estuvo determinada principalmente por la estimulación de los procesos de crecimiento y la reducción de las defensas químicas, constituidas por compuestos de sabores desagradables. También se sugiere que en una segunda fase los atributos asociados a la palatabilidad de las mandarinas, especialmente la acidez, se mejoraron progresivamente a través de modificaciones específicas. Otros cambios de relevancia se relacionaron con la regulación de genes involucrados tanto en la síntesis de sustancias básicas de agradable aroma y sabor, como en la modulación de transportadores de azúcares. Por lo tanto, la transición entre las papedas incomestibles y las mandarinas de tipo ácido se caracterizó esencialmente por una drástica reprogramación de la expresión génica de rutas metabólicas fundamentales, mientras que las mandarinas modernas evolucionaron posteriormente mediante un refinamiento progresivo de las propiedades relacionadas con la palatabilidad. En su conjunto, estas observaciones sugieren que en cítricos, las tasas de mutación de los tejidos somáticos son consistentes con la idea de que las mutaciones desempeñaron un papel importante en las últimas fases de la domesticación.

Resum

La domesticació dels cítrics és un procés en gran part desconegut. Les investigacions poratades a terme fins al moment indiquen que les varietats de cítrics comercialitzades hui en dia són, en general, el producte de introgressions ancestrals intraespecífiques i interespecífiques. Les introgressions de tipus intraespecífic van tindre lloc entre les poblacions de dues subespècies antigues de mandarina tot aportant el comportament apomíctic que la majoria de les mandarines actuals i altres cítrics relacionats encara conserven. Una vegada que aquestes barreges genètiques o *admixtures* es consolidaren, diferents introgressions interespecífiques de *pummelo* establiren noves característiques en el genoma de mandarina. Durant el procés de domesticació, la variabilitat genètica dels cítrics va ser assolida mitjançant mutació espontània i, per mitjà de la pràctica de l'empelt, genotips sense llavors es van estendre ràpidamente, accentuant el paper de les mutacions en la selecció de cítrics com les taronges i les mandarines modernes.

En aquest treball, mostrem com les mutacions somàtiques en cítrics es propaguen seguint un patró iteratiu determinat per un model de ramificació simpodial i, en conseqüència, queden agrupades al llarg de l'arbre, on algunes d'aquestes mutacions romanen fixades. En aquest escenari, un arbre pot ser considerat un mosaic genèticament format per diferents genomes, en el qual les branques més joves acumulen un nombre major de mutacions. De mitjana, el nostre arbre experimental de Clementina mostra una ràtio de mutació de 4.4×10^{-10} bp^{-1} yr^{-1}, arribant a incloure en total entre 1500 i 5000 variants i produint-hi 1 mutació somática en cada meristem axil·lar. Aquest nombre relativament elevat de mutacions està en línia amb el gran nombre de varietats provinents de mutacions espontànies que es comercialitzen en cítrics.

Per tal d'abordar la identificació de trets essencials adquirits durant la domesticació, s'ha analitzat el transcriptoma de fruits en desenvolupament de la varietat silvestre i incomestible papeda d'Ichang, la mandarina àcida Sun Chu Sha Kat i tres segregants comestibles derivats d'un creuament entre mandarines comercials. Basant-nos en els resultats obtinguts, s'ha proposat que durant la transició de les papedes incomestibles a les mandariens àcides, la domesticació es va caracteritzar per una primera fase de canvis radicals en l'expressió dels gens que regulen rutes fonamentals del metabolisme primari i secundari. Aquesta fase va estar determinada principalment per l'estimulació dels processos de creixement i la reducció de les defenses químiques de gust desplaent.

També es suggereix que en una segona fase atributs associats a la palatabilitat de les mandarines, especialment l'acidesa, van ser millorats progressivament mitjançant modificacions específiques. Altres canvis de rellevància van estar relacionats amb l'estimulació de gens involucrats tant en la síntesi de substàncies bàsiques d'agradable aroma i sabor, com en la modulació de transportadors de sucres. Per tant, la transició entre les papedes incomestibles i les mandarines àcides va estar caracteritzada essencialment per una dràstica reprogramació de l'expressió gènica de rutes metabòliques fonamentals mentre que les mandarines modernes evolucionaren posteriorment mitjançant un refinament progressiu de propietats relacionades amb la palatabilitat. En conjunt, aquestes observacions suggereixen que, en els cítrics, les ràtios de mutació dels teixits somàtics són consistents amb la idea que les mutacions tingueren un paper important en les últimes fases de la domesticació.

Acknowledgements

Es difícil delinear con nombres propios un plural tan amplio. Difícil y arriesgado. Por eso, seré breve.

En primer lugar, he de destacar el papel indispensable de mis compañeros del Centro de Genómica del IVIA. Todos ellos han posibilitado de algún modo este trabajo. No hace falta nombraros, sabéis quien sois.

De manera especial, sin embargo, quiero reconocer como, desde el primer momento, la respuesta a todas mis dudas, inseguridades y falta de criterio la he encontrado en ti. Has sido el ejemplo que mi actitud necesitaba. Sin más, gracias, Carles.

Del mismo modo, debo admitir que adentrarse en el mundo científico guiada por la competencia del apellido Talón ha sido un honor a todos los niveles. Cada palabra de este trabajo es garante de ese hecho. Gracias a ti también, Manolo.

I, siga com siga, aquestes paraules no poden obviar altres certeses que, si bé resulten menys pragmàtiques, esdevenen igualment importants. Per això, he de confessar que ha sigut alliberador comprovar com l'amistat no entén de condicions. Amb un significat que- sortosament per a mi- desmereix el de cada errada que l'amenaça, avançar ha sigut l'única possibilitat.

En aquest sentit, el triomf personal més valuós que he assolit ha estat trobar en els anys, els bacs i l'experiència viscuda durant aquesta etapa l'oportunitat de reafirmar-me en allò que sempre he sospitat. I, és que, al remat, la conclusió no pot ser diferent: la meua sort porta el vostre nom. Gràcies, família.

List of Figures

Introduction

Chapter 1

Chapter 2

Discussion

List of Tables

Chapter 1

Chapter 2

Abbreviations

CLM: Clementine

DEG: Differential Expressed Gene

DNA: Deoxyribonucleic Acid

ICH: Ichang Papeda

INDEL: Small Insertion and Deletion

MITE: Miniature Inverted-repeat Transposable Element

RNA: Ribonucleic Acid

S1, S2, S3: Segregant 1, Segregant 2, Segregant 3

SNP: Single-Nucleotide Polymorphisms

SCM: Sun Chu Sha Kat Mandarin

WGS: Whole Genome Sequencing

WMU: W. Murcott

INTRODUCTION

Introduction

1. Citrus Production

Citrus is one of the most important fruit crops worldwide and its vast cultivation area covers a broad belt around temperate and tropical countries. The complete range of edible citrus varieties, which are produced in warm regions located mainly in China, Southeast Asia, the Mediterranean Basin, the Caribbean Sea, California, South of Africa, and several regions of South America, represents 17.9% of total fruit crop productivity. Among the fruit products, sweet oranges, mandarins, grapefruits and lemons hold the greatest importance in the industry. For example, in 2020, sweet orange and mandarin production reached 75.5 and 38.6 million tons, respectively (FAO, 2022). Notably, production, consumption and trade of mandarins have doubled its economic relevance along the last couple of decades (Spreen et al., 2020). In domestic terms, Spain is the largest citrus producer in the European Union and sixth in the world, with an annual production of more than 6.8 million tons. Stats show that Spain is the world major exporter of fresh fruit, with figures verging on the 4 million tons (EUROSTAT Comext, 2021). Furthermore, the Valencian Community is the main producing region, in terms of land used (near 55%) and total production (more than 3 million tons per year, almost 50% of national production) (MAPA, 2022). One important aspect concerning trade is the way citrus are marketed. Citrus fruits are sold as fresh or processed products (fresh and canned juices), and as by-products from juice processing. In either way they are consumed, citrus products serve as a daily nutritional intake since they constitute a rich source of vitamins, minerals and dietary fiber which are essential on nutritious and tasty diets. There are other biologically active, non-nutrient compounds (mostly derived from secondary metabolism as limonoids, anthocyanins, essential oils or carotenoids) that may have beneficial effects on human health (Lv et al., 2015).

Although the importance of this crop in a global context, is beyond question, there are several challenges that currently are threatening its viability. In recent years, the globalization of citrus production, the spread of devastating diseases and the consequences of climate change are menacing the future of the citrus industry. Studies related to global warming escalation, for instance, foresee an increase in temperatures coupled with a decrease in precipitation which will cause an intensification of drought

periods and a reduction in the amount of water devoted for irrigation. In this scenario, the survival of this crop depends critically on the generation of new genetically superior cultivars of higher quality and productivity, with longer marketing periods and resistant to both abiotic (salinity and drought) and biotic (pests and exotic diseases) stresses. However, traditional breeding programs have been proved to be rather limited because of the occurrence of several constrictions and limitations of the citrus biology that hinder their success. Citrus breeders face many difficulties inherent to a crop that shows asexual reproduction, and whose life cycle may include parthenocarpy, auto-incompatibility, sterility, polyembryony, apomixis and long periods of juvenility. Furthermore, most commercial varieties are derived from spontaneous mutations clonally propagated, a circumstance that narrows the genetic diversity available for the development of new cultivars with proper and desirable agronomical traits.

2. Citrus Genomics

For the last twenty years, genetic and genomic advances have been providing researchers with new tools and approaches to tackle citrus breeding. The rise of genomics has enabled to gain profound knowledge related to biological and evolutionary questions. This circumstance is speeding up the establishment of new breeding techniques. The *de novo* assembly of sweet orange (*Citrus sinensis* cv. Valencia) draft (Xu et al., 2013) and the publication of the high-quality reference genome derived from a haploid Clementine (*C.* × *clementina* cv. Clemenules) (Wu et al., 2014) were the turning point for the era of citrus genomics. Both *C. sinensis* and *C. clementina* genomes and those obtained thereafter for citron (*Citrus medica* L.), pummelo (*Citrus maxima* (Burm.) Merr.), Ichang papeda (*Citrus cavaleriei* Swingle, 1913 H.Lev. ex Cavalerie, 1911, or *Citrus ichangensis* Swingle, 1913) (Wang et al., 2017*b*), mandarin (*Citrus reticulata* hort. ex Tanaka Blanco) (Wang et al., 2018) and kumquat (*Fortunella hindsii* (Champ. ex Benth.) Swingle) (Zhu et al., 2019), granted the access to the vast nucleotide diversity defining this genus, opening major opportunities to elucidate the structure of the citrus genome. Citrus species show a well-conserved genomic structure in terms of genome size, gene number and synteny. The genomes in the genus *Citrus* roughly comprises 300 Mb of sequence, that are arranged in nine true chromosomes (2n = 18) where about 27,000 genes have been annotated (National Center for Biotechnology Information, 1988).

The analysis of single-nucleotide polymorphisms (SNPs) not only enabled quantification of genetic differentiation between diploid genomes but also let to the generation of a scale of divergence between them as an indication of ancient speciation events. In addition, the determination of the degree of heterozygosity, among other parameters, helped to discriminate among citrus *pure species*, i.e., pure genomes, hybrids, and genetic admixtures. These approaches identified up to ten basal citrus species and a range of hybrids and admixtures showing different degrees of introgression, mostly of pummelo (*C. maxima*) (Wu et al., 2018, 2021).

In addition to SNPs, other genetic variants have been used for surveying the extant variability, as for example, that of citrus chloroplast haplotypes, whose phylogenetic relationships were determined using small insertions and deletions (INDELs) (Carbonell-Caballero et al., 2015; Maddi et al., 2018). The use of structural variations due to major chromosome alterations, on the other hand, are currently used as molecular markers for cultivar identification. In citrus, as in other crops, the identification of large chromosomic rearrangement and mobile element insertions (Terol et al., 2015; Borredá et al., 2019) have become a preferential matter of study, since these alterations may show causative relations with agronomical traits (Gabur et al., 2019). Finally, studies based on WGS are paving the way for untangling evolutionary citrus history.

3. Citrus Transcriptomics

RNA-seq technique is part and parcel of the advent of genomic approaches that have been driving citrus research during the last decade. The majority of these transcriptomic studies are focused on the characterization of physiological aspects related to fruit quality and yield. In particular, the identification of genes connected to attributes impacting on consumer's opinion, such as taste or appearance, are long pursued. For example, Hussain et al., (2020), delved into the expression profile of a *Citrus sinensis* mutant with low sugar content, while Zhang et al., (2021), proposed a list of 24 transcription factors which might play important roles in carotenoid and flavonoid accumulation in another red-flesh mutant of Navel orange. The ripening process also constitute a highly interesting matter as proved by the publication of different research articles (Wu et al., 2016; Wang et al., 2017*a*). Expanding the scope of that subject, in a recent paper, Borredá et al., (2022), used RNA sequencing data to study the transcriptomic profiles of wild and commercial varieties during early stages of fruit

ripening. In this case, the authors approached the analysis of the results under the point of view of domestication.

Tolerance mechanisms in response to biotic and abiotic stresses have also been extensively investigated through transcriptome analysis. For instance, the effects of either natural or induced viral infections in citrus varieties have often been studied in comparative gene expression analysis (Fu et al., 2017; Ramirez-Pool et al., 2022). Furthermore, the role of citrus rootstocks in the response of the scions to the environmental conditions has also been matter of study, as reported in Primo-Capella et al., (2022), that analyzed the modulating effects that rootstocks exert in cold responses.

Although experiments that use fruit tissues and, especially, those centered on sweet orange, are numerous, RNA-seq data have also been used to screen citrus transcriptome landscapes of vegetative and reproductive structures from several species spanning different phylogenetic groups of the genus *Citrus* (Terol et al., 2016).

4. Citrus Origin

Citrus species are classified in the Geraniales/Sapindales Order, the Rutaceae Family and the Aurantioideae Subfamily. The Aurantioideae group is itself subdivided into two tribes known as Clauseneae and Citreae. This last tribe comprises the subtribe Citrinae which include 13 genera. Among these 13 genera, six of them, *Fortunella*, *Eremocitrus*, *Poncirus*, *Clymenia*, *Microcitrus* and *Citrus* are regarded as 'true citrus fruit trees' according to Swingle and Reece (1967). Members of *Citrus* bear fruits that are botanically considered hesperidia. Specifically, they are modified berries in which the fleshy parts are divided into segments. Those segments are, in turn, filled with elongated sacs in which thin-walled cells contain a watery juice (Tadeo et el., 2020). Cultivated citrus in general are evergreen shrubs or small trees, but wild species can be notable taller and bigger.

Attempts to classify citrus based on morphological and anatomical characteristics have generally been unproductive, because domestication has produced thousands of types and varieties of citrus that as we know today, are genetic admixtures of basal pure species combining mixtures of traits artificially selected. Widespread interspecific hybridization and the long and complex history of backcrossing and clonal propagation that took place during citrus domestication, have produced so many closely related or similar varieties, that it is extremely difficult to meaningfully classify them. Renowned

citrus botanists, for instance, have repeatedly incurred in great disagreements. This lack of consensus arose from difficulties in setting species boundaries using only phenotypical perspectives. For instance, the large group of 166 different citrus species determined by Tanaka (1954) was reduced to 16 by Swingle (Swingle and Reece, 1967). Tanaka also named more than 30 species of mandarins while Swingle only recognized 3 species. Likewise, establishing the geographical origin of citrus was a question casting certain controversy. Notwithstanding, the irruption of genomics as stated before, is currently providing a broad amount of original information in this field. This circumstance is particularly impacting the notions about citrus origin, taxonomy, and genealogy (Xu et al., 2013; Wang et al., 2018; Wu et al., 2014, 2018, 2021) as new knowledge is laying the groundwork for our current understanding of the citrus concept. The distinction between pure species and admixed genomes provided by WGS data and the combined analysis of genetic divergence and heterozygosity constitute the premises on which solid taxonomic and phylogenetic hypothesis have been proposed. In a comparative genomic analysis Wu et al., (2018) identified 10 progenitor citrus species clearly delimiting the crucial events which drove speciation.

Under this new framework, the discovery of a Late Miocene leaf fossil named as *Citrus linczangensis* in the Chinese province of Yunnan, was determinant for dating the phylogenetic citrus tree. This paleontological record exhibited several traits common to different citrus clades which justified its consideration as ancestor of the whole genus (Xie et al., 2013). The phylogenetic analysis revealed that seven out of the 10 ancestral species evolved in a first phase of the speciation process during the late Miocene in East and Southeast Asia. According to Wu et al., (2018), *C. medica* , *C. maxima* , *C. reticulata* , *C. micrantha*, *C. ichangensis*, *F. japonica* and *C. mangshanensis* emerged in an area recognized as a diversification center, comprising the southeast foothills of the Himalayas, and delimited by eastern Assam, northern Myanmar and western Yunnan. Citrus speciation was coincided with the onset of a period of global cooling and declining of CO_2 levels, conditions that led to a dramatic climatic transition in Southeast Asia, from a monsoon-dominated model to an environment characterized by aridity and seasonal droughts. In Wu et al., (2021), the basic taxon of *C. reticulata* was further investigated and the authors proposed the existence of two mandarin subspecies. Accessions traditionally known as *common mandarins* compose one of these groups, while the other, defined by *mangshanyeju* haplotypes, includes Chinese related wild

mandarin populations. The diversification event was situated approximately 1.5 million years ago. The other three species, *C. glauca* (synonym of *Eremocitrus glauca*; australian desert-lime), *C. australis* (synonym of *Microcitrus australis*; australian round-lime) and *C. australasica* (synonym of *Microcitrus australasica*; australian finger-lime), probably arose in a second phase of speciation during the early Pliocene in Oceania.

5. Citrus Somatic Mutation

Somatic mutations are changes in DNA sequence that arise as genetic events affecting only one dividing cell during its mitotic cycle. However, its occurrence is very relevant because may constitute a potential start of a new cell lineage. In general, the degree of structural modification they caused in the chromosomes, determines their effects in terms of genetic expression and, thus, of individual fitness. Somatic mutations, therefore, lie at the heart of the genotypic variation that brings about both gain and loss of protein functions. Accordingly, one of the potential effects of somatic mutations is the generation of variability which regularly contributes to plant evolution. The accumulation of genetic variation is, indeed, a process that differentiates individuals as they change the characteristics of populations by recombination, natural selection, and drift. In the long term, genetic variation acts driving speciation and the diversification of lineages.

DNA strands can be physically altered as a result of errors during DNA replication or through lesions acquired by exposure to environmental mutagens. The extent of the alterations undergone by DNA molecules ranges from a single site change, i.e., point mutations that lead to single-nucleotide polymorphisms (SNPs), small insertions and deletions (INDELs), and chromosomic rearrangements, which involve millions of nucleotides (structural variants). Somatic mutations are very often the result of the activity of mobile elements.

During the last decade, multiple studies characterizing somatic mutation incidence in citrus genomes have been performed, since polymorphism information is ready accessible mapping the resequencing data against reference genomes. For instance, Caruso et al., (2018), carried out research detecting natural and induced mutations and generating a robust and reliable set of sequence variations in sweet orange. In a crop clonally propagated like that, intraspecific diversity is truly convenient for

fingerprinting and varietal traceability. Also, Wang et al., (2021), have recently published a complete survey on the somatic diversity of 114 *Citrus sinensis* mutants, including SNPs, structural variations, and transposable elements. In an alternative approach, research has been focused on the association between mutations and particular phenotypical features. Terol et. al., (2015, 2019), for example, characterized the structural consequences of a Mutator-like element activation comparing the genomes of bud sport mutants of Clementine. In these works, a MAD box transcription factor located in a 2 Mb hemizygous deletion present in one of the mutants was associated with the earliness-ripening trait.

The rates at which single-nucleotide variants arise impact on multiple biological aspects. Data substantiate that mutation rates, that are calculated in a per-time basis- that is, per year or per generation- vary almost 1,000-fold between the different species (Figure 1).

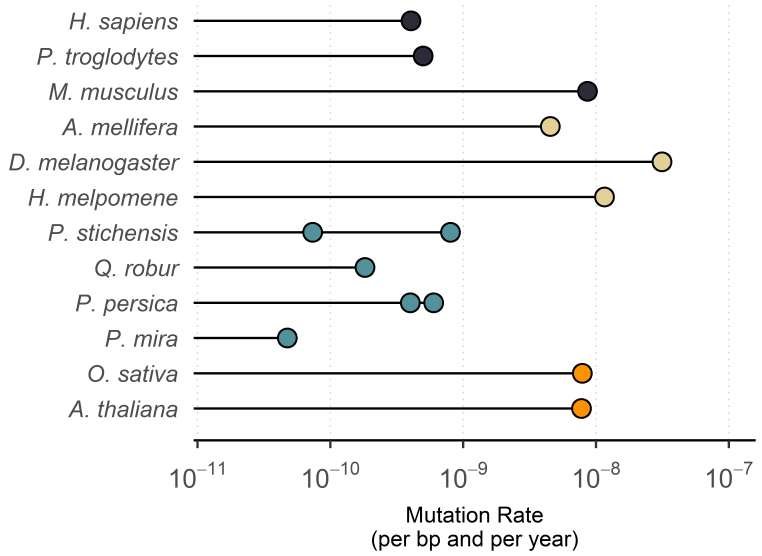


Figure 1. Somatic mutation rates reported for different animal and plant species. Orange: Annual plants (Ossowski et al., 2010 and Wang et al., 2019); Turquoise: Perennial plants (Xie et al., 2016; Schmid-Siegert et al., 2017 and De La Torre et al., 2017); Yellow: Insects (Keightley et al., 2015; Schrider et al., 2013 and Yang et al., 2015); Black: Mammals (Uchimura et al., 2015; Venn et al., 2014 and Kong et al., 2012). Data interpreted as in Hanlon et al., 2019.

Base-substitution mutation rates in eukaryotes, for instance, range from ~5 × 10^{-11} to 5 × 10^{-8} per site per year (Lynch, 2010). In plants, mutation rates calculated in long-lived organisms are seemingly lower than those found in herbaceous plants (Smith and

Donoghue, 2008), as comparing, for example, woody plants with well-studied annual species such as rice (*Oryza sativa* L., 1753) or Arabidopsis (*Arabidopsis thaliana* (L.) Heynh., 1842). As reported by Xie et al., (2016), on a per-year scale perennial plants evolved at a reduce pace compared to short-lived annual.

6. Citrus Genetic Admixtures

As mentioned above, based on the analysis of different parameters, including the study of local ancestry across different genomic regions, Wu et al., (2018), proposed the existence of pure genomes, hybrids and genetic admixtures in the genus *Citrus*. Pure genomes show low intraspecific diversity (0.1-0.6%) in comparison with admixtured citrus, composed of two or more species that exhibit higher segmental heterozygosity (1.5-2.4%). Bimodal distributions are observed in these citrus accessions that derived from complex backcross mechanisms (Figure 2A).

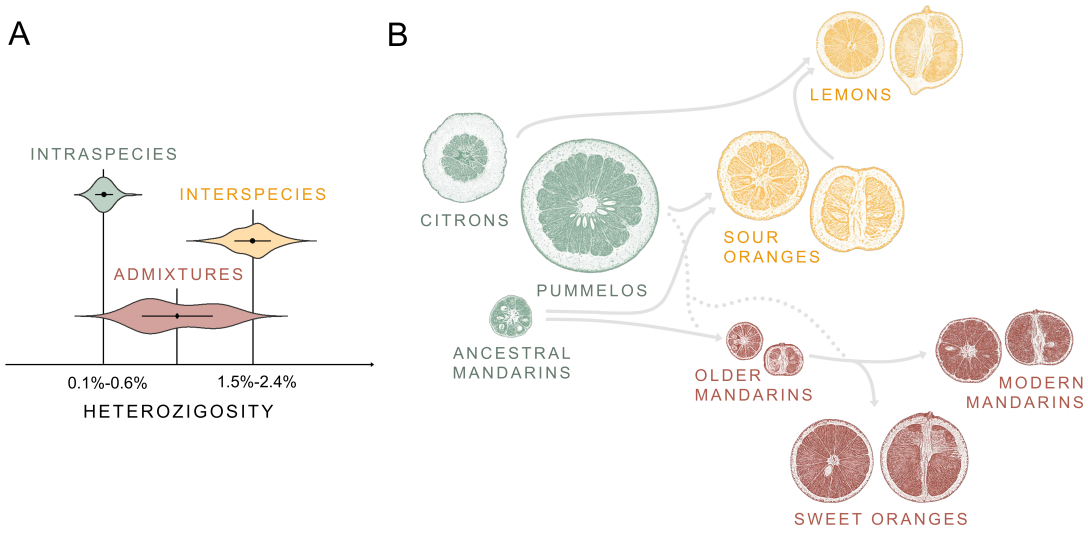


Figure 2. Pure genomes, hybrids and genetic admixtures of citrus as related to domestication. **(A)** Typical heterozygosity values in citrus genomes. Green: pure species; Yellow: hybrids; Indian red: genetic admixtures. **(B)** Genealogy of major commercial citrus varieties. Green: parental pure species; Yellow: acidic varieties; Indian red: edible varieties. Adapted from Wu et al., (2014, 2018, 2021).

These authors reported that three of these pure species, namely, citrons (*C. medica*), mandarins (*C. reticulata*) and pummelos (*C. maxima*) were the foundational species of the cultivated citrus since the genomes of the commercial varieties, including mandarins, lemons, limes, oranges and grapefruits, contain at least two of these three genomes (Figure 2B). Commercial mandarins, for instance, are mainly characterized by

intraspecific introgression of *mangshanyeju*, that contributed the apomictic behavior and by pummelo interspecific introgressions. Oranges are probably the result of at least one additional backcross with the maternal pummelo. Mandarins can be classified, therefore, according to the amount of pummelo introgression. Apart from the pure mandarins, i.e., pure genomes, that do not carry pummelo traces, a first group of mandarins is formed by those having a small proportion (1-10%) of *C. maxima* genome. The proportion of pummelo alleles for the second group is distinctly higher (12-38%) as they span longer segments.

That classification appears to mirror the sequential stages leading to the admixture patterns observed today. The data suggest the occurrence of an initial phase in which just one pummelo tree was involved, since the first group of mandarins display two common *C. maxima* haplotypes, likely incorporated into the maternal *C. reticulata* genome in an early introgression episode. Then, in just about five or six consecutive backcrossing events that shattered those pummelo segments, the set of older and traditional mandarins were probably originated. Sweet orange and modern mandarins, on the other hand, appear to be the result of additional pummelo introgressions.

These observations indicate that the term *mandarin* as it is popularly used, encompasses a set of different genotypes including pure and wild species bearing inedible fruits and palatable commercial accessions with genomes admixed in different proportions.

The finding that pummelo introgression was a general feature of all edible cultivars (Wu et al., 2014, 2018), points out that those introgressed fragments may include crucial traits for domestication and, hence, they were likely selected and fixed during the initial stages of that process.

7. Citrus Domestication

Domestication is a co-evolutionary process that occurs when wild plants are brought into cultivation by humans originating new species. It is currently accepted that plant domestication started at the outset of the Holocene, approximately 12,000 years ago, when hunter-gatherer societies began growing plants as sources of food and fiber (see cites in Purugganan, 2019). It is estimated that up to 2,500 plant species were domesticated, as a result of this lifestyle transition, the so-called *Neolithic revolution*, which occurred independently in different regions around the world (Meyer et al., 2012).

The impact of this human-assisted selection is imprinted in the genomes of domesticated crops, as they diverged clearly from those of wild relatives. Fixation of beneficial mutations by means of positive selection invariably causes a severe reduction in genetic diversity, generating genome sweeps that include genetic hitchhiking effects. The evolution of particular genome areas under selection pressures departs from neutral model (Gonzalez-Ibeas et al., 2021*a*, *b*). The ongoing nature of this evolution process generates a morphological and genetic continuum in which artificial pressures cause cultivated populations to eventually diverge from their wild progenitors (Miller and Gross, 2011). Domestication syndrome is defined as the suite of traits that marks a crop's divergence from its wild ancestor. The comparison of diverse crops has evidenced the convergence nature of domestication since different traits have been selected recurrently in different species. Most prevailing features in plant domestication are derived from changes in secondary metabolites which affect flavor, pigments, and toxicity, although reproductive characteristics such as homogeneity of flowering and ripening times or the development of non-shattering seeds in cereals are also relevant targets for domestication (cites in Meyer et al., 2012)

Tree crop domestication progresses at a pace that is markedly determined by the long generation times characteristic of woody plants. In contrast to short-lived crops and grains, whose life cycles allow many selection rounds, the vegetative and reproductive biology of woody perennials has burdened the fixation of desirable alleles. In these plants, long juvenility phases, mechanisms to avoid selfing, high rates of inter- and intraspecific hybridization, extensive population genetic variation, and limited population structure are pivotal features influencing the way they evolve in nature and under domestication. In this scenario, the changes in reproduction biology are the factor determining the outcomes of each domestication process (Miller and Gross, 2011).

Citrus domestication is characterized by a complex history of intra- and interspecific admixtures and, therefore, current diversity is intimately linked to the widespread of ancient introgressions. Asexual reproduction and propagation, first by means of nucellar polyembryony and later by grafting of selected individuals, is another undeniable determinant factor in the process of domestication of citrus. Spontaneous mutations also provided further diversity while selection of accessions arisen from natural hybridization, generated the extensive network of relatedness describing current citrus cultivars (Wu et al., 2018). Domestication targets in fruit-bearing crops are commonly

associated to production and fruit quality. In citrus, desired domesticated characteristics were seedlessness, yield, flavor, juiciness, acidity, peel color, fruit growth, and ripening among the most important. For instance, color peel, a characteristic that increases fruit look appealing, responds in citrus mostly to the accumulation of different type of carotenoids. Rind red pigmentation, in fact, is mainly caused by the presence of C30 apocarotenoids. The wide spectrum of colors displayed by commercial varieties appear to be linked to the occurrence of different mutations affecting the metabolism of these isoprene-derived compounds. Carotenoid cleavage dioxygenase encoded by the CCD4b gene (Zheng et al., 2019) together with the expression of two β-lycopene cyclases (Alquézar et al., 2009; Zhang et al., 2012) in ripening citrus fruits are reported to be key genes regulating the accumulation of downstream pigments. Other components contributing to coloration are anthocyanins. They are pigment compounds that are extensively found in wild citrus accessions while lacking in most cultivated ones, especially mandarins. Their presence, however, in blood oranges constitutes a classic example of artificial selection. It has been reported that a mutation affecting the promoter sequence of a MYB transcription factor known as *Ruby* is the genetic reason causing the colored phenotype of those oranges (Butelli et al., 2012, 2017).

Acidity ranks at the top of the traits determining citrus flavor quality and consequently, was one of the main targets pursued by early breeders and producers. Wild mandarins bear, in general, extremely acidic and unpalatable fruits, while a vast assortment of extant mandarin varieties with phenotypes ranging from sour to non-sour tastes can be accounted for. The biochemical factor governing the acidic taste of citrus flesh is the pH reached in the vacuoles of juice sacs. Extremely low values, for example, are detected in certain citrus types, like citrons, lemons, and limes (Hussain et al., 2017). In cells of the juice sacs, the vacuolar high concentration of free H^+ ions is buffered by citrate that possess a relatively high pH-buffering capacity (Shimada et al., 2006). The steep pH gradient caused by protons allows for the import and sequestration of this organic acid through the tonoplast into the vacuolar lumen. Genes related to the metabolism of organic acids have been largely surveyed as they very likely play a role in acid content reduction, such as cytosolic aconitase (ACO; Terol et al., 2010), NADP-isocitrate dehydrogenase (Wang et al., 2018, Wu et al., 2018), or glutamate decarboxylase (GAD) (Liu et al., 2014). Besides, activity of the GABA shunt contributes to the fluctuation of citric acid concentration (Cercós et al., 2006, Sheng et al., 2017). The relevance of

vacuolar transport in pH stability has compelled to study also transport mechanism across tonoplast. In an elegant publication, Strazzer et al., (2019), reported on the major components of the complex regulating hyperacidifycation of vacuoles in juice sacs of acidic citrus. These authors highlighted the role of two interacting P-ATPases, namely, CitPH1 and CitPH5, which expression levels correlated with acidity in acidic and acidless citrus varieties. Unrelated mutation events in distinct regulator genes, that likely arose independently multiple times, are aimed to be the causative genetic changes of the trait. Among those key regulators, there are two transcription factors of the WRKY and MYB families, CitPH3 and CitPH4, respectively, a WD-repeat protein, CitAN11, and a helix–loop–helix protein ANTHOCYANIN1, CitAN1. The latter, also known as *Noemi¸* acts modulating the expression of the above-mentioned gene *Ruby* (Butelli et al., 2019).

In addition to these examples, there are other predominant traits selected during domestication. One of these characteristics of substantial importance is derived from the natural asexual propagation that apomictic seed dispersal allows. In *Citrus*, apomixis is generally, but not always, manifested by polyembriony. Apomixis was acquired through the mutation of CitRKD1 gene (Shimada et al., 2018; Wang et al., 2017*b*) in *mangshanyeju* ancestral mandarins (Wu et al., 2021) and thereafter was propagated and fixed in other wild and domesticated citrus types. Apomixis refers to the development of embryos from the somatic nucellar cells that surround the fertilized egg in the seed. In terms of reproduction, it implies the possibility of generating an offspring genetically identical to the mother parental preventing the segregation of specific alleles and, therefore, constitutes a natural mechanism to preserve a desired phenotype over time.

OBJECTIVES

Objectives

The main goal of this work was to contribute to the understanding of the citrus domestication process, in order to identify important genes of agronomic relevance and organoleptic value. To this end, two main objectives related to the genomics of citrus genetic admixtures, which constitute the bulk of current commercial citrus varieties, have been addressed.

1. The first objective was designed to elucidate the dynamics of single-nucleotide polymorphisms, and to determine absolute and relative rates of spontaneous mutations in Clementine mandarin, a globally accepted commercial citrus genetic admixture.

2. In a second objective, the work was focused on the study of the differences in gene expression in developing fruits, between non-edible wild species and commercial genetic admixtures of citrus.

CHAPTER 1

CHAPTER 1

Single-nucleotide Mosaicism in Citrus: Estimations of Somatic Mutation Rates and Total Number of Variants

Perez-Roman, E., Borredá, C., López-García Usach, A., & Talon, M. (2022). Single-nucleotide mosaicism in citrus: Estimations of somatic mutation rates and total number of variants. *Plant Genome*, 15:e20162. https://doi.org/10.1002/tpg2.20162

The student contributed to this article:

1. Curating WGS data, performing bioinformatics analysis and conducting experimental validations.
2. Creating and preparing visual items: plotting results, designing figures and arranging tables.
3. Reviewing and editing the original draft.

Abstract

Most of the hundreds of citrus varieties are derived from spontaneous mutations. We characterized the dynamics of single-nucleotide mosaicism in a 36-yr-old clementine (*Citrus ×clementina* hort. ex Tanaka) tree, a commercial citrus whose vegetative behavior is known in detail. Whole-genome sequencing identified 73 reliable somatic mutations, 48% of which were transitions from G/C to A/T, suggesting ultraviolet (UV) exposure as mutagen. The mutations accumulated in sectorized areas of the tree in a nested hierarchy determined by the branching pattern, although some variants detected in the basal parts were also found in the new growth and were fixed in some branches and leaves of much younger age. The estimate of mutation rates in our tree was 4.4×10^{-10} bp^{-1} yr^{-1}, a rate in the range reported in other perennials. Assuming a perfect configuration and taking advantage of previous counts on the number of total leaves of typical clementine trees, these mutation determinations allowed to estimate for the first time the total number of variants present in a standard adult tree (1,500–5,000) and the somatic mutations generated in a typical leaf flush (0.92–1.19). From an evolutionary standpoint, the sectoral distribution of somatic mutations and the habit of periodic foliar renewal of long-lived plants appear to increase genetic heterogeneity and, therefore, the adaptive role of somatic mutations reducing the mutational load and providing fitness benefits.

Core Ideas

- Variant distribution follows an iterative process according to the branching pattern.
- Variants accumulate with time, the youngest the branch the highest the mutation rate.
- Only a few older variants are fixed in most leaf flushes regardless of their ages.
- Clementine trees carry 1,500–5,000 mutations, and each leaf flush generates a new variant.
- A tree is a mosaic genetically composed of different genomes distributed by sectors.

Introduction

Genetic mosaicism generally arises from somatic mutations that occur during mitosis and are passed on to descendant cells. Although theoretical models initially predicted that mosaicism should be rare in eukaryotes, it is currently accepted that animal and plants are, in fact, genetic mosaics composed of many different cell populations (Gill et al., 1995; Lupski, 2013). Plants become genetically mosaic as they accumulate spontaneous mutations that occur at any stage of development in both undifferentiated and differentiated cells. It has been known for a long time, for instance, that plant chimeras are mosaics composed of distinct genotypes (Szymkowiak and Sussex, 1996).

Somatic mutations involve genetic changes mostly occurring during chromosome segregation or DNA replication, leading to structural genomic rearrangements and single-nucleotide variation among other alterations (Lupski, 2013). While there is an increasing number of reports describing the relevance of chromosomal rearrangements, for instance, on the generation of crop varieties (Terol et al., 2015; Alonge et al., 2020) and hence on plant evolution, the evolutionary meaning of single-nucleotide polymorphism (SNP) accumulation remains unresolved. A major argument supporting the role of single-nucleotide mosaicism in plants is based on the idea that intraorganismal heterogeneity, especially in long-lived plants, may contribute to plant defense, at least partially, increasing resistance or tolerance to short-lived pests and pathogens (Gill et al., 1995). The underlying assumption of this hypothesis implies that the variation in somatic mutations emerging from different meristems may alter susceptibility to insect attacks and therefore may confer herbivore and pathogen resistance to individual plants or individual branches in long-lived trees (Simberloff and Leppanen, 2019). Somatic mutations do not only accumulate in a certain plant, as they can also be transmitted to the gametes and hence to the offspring as recently reported (Plomion et al., 2018; Wang et al., 2019), contributing in this way to the genetic variability of many traits. Several studies have demonstrated that the branches of a single tree may show significant variation in fitness traits (see citations in Schoen and Schultz, 2019). However, the relative long lifespan of a tree should also allow accumulation of harmful somatic mutations that after passing to the offspring may lead to a phenomenon of ‘mutational meltdown’ resulting in the loss of fitness. Thus, the dynamics of somatic mutations, their general functions and mechanisms of regulation, and therefore, their roles during evolution are poorly understood.

Among the pivotal determinants of the somatic mutation buildup are the number and patterns of stem cell division, events that eventually determine tree architecture (Burian et al., 2016; Schoen and Schultz, 2019). Tree architecture, in turn, depends upon many anatomical and physiological factors including the arrangement and organization of the apical and axillary meristems. In addition, plant meristems are stratified structures generally consisting of three different cell layers, as those found in citrus (Frost and Krug, 1942), giving rise to separate tissues. These layers may accumulate mutations independently generating clonally diverse cell lineages that form chimeric organs with distinct proportions of the mutated alleles, a circumstance that may hamper the determination of mutations rates. The activity of the plant meristems determines the patterns of iterative branching that in long-lived plants may result in a high number of branches and literally thousands of terminal twigs. In the sympodial pattern of branching, as in citrus, the shoot apex growth is generally arrested, while axillary meristems initiated in the axils of leaves give rise to new shoot branches, providing to the plant an unlimited potential for growth. In contrast, in the monopodial model as in the conifers, a unique apical meristem gives rise to the new branches. In most angiosperms, therefore, the number of cell divisions contributing to mutation accumulation is delimited by the number of cell divisions occurring between the apical and an axillary meristem and the number of branching points of the branches (Burian et al., 2016).

The determination of the mutation rates, however, is time-consuming, laborious, and technically challenging. Mutation rates are calculated in a per-time basis, that is, per year or per generation. Based on mutations accumulated during divergence between monocots and dicots, Wolfe et al. (1987) initially estimated that in annual plants, the rates of substitutions per site range from 5.0×10^{-9} to 3.0×10^{-8}. In long-lived perennials, it has been reported that these rates might be as high as 25 times those of short-lived annuals (Klekowski and Godfrey, 1989). Genome analyses have provided estimations that, in general, agree with these tendencies. Ossowski et al. (2010), for instance, reported that mutations rates in *Arabidopsis thaliana* (L.) Heynh. are in the order of 7.0×10^{-9} substitutions per site per generation, while perennials, however, tend to present higher rates of fixed mutations, as reported in the Napoleon Oak, between 4.2 and 5.2×10^{-8} substitutions per site (Schmid-Siegert et al., 2017) or 2.7×10^{-8} in Sitka spruce (*Picea sitchensis* (Bong.) Carr.) (Hanlon et al., 2019). Since perennials are long-

lived organisms, the per-year mutation rates are generally lower than those exhibited by herbaceous taxa (Xie et al., 2016). Thus, in Sitka spruce, the per-year base substitution rate estimations range from 1.2×10^{-10} to 5.3×10^{-11} (Hanlon et al., 2019). In a comprehensive work comparing several perennials and annuals, Wang et al. (2019) reported that on a per-year scale, long-lived perennials apparently evolved slower than short-lived annuals since those generally exhibited lower mutation rates (10^{-10}–10^{-9}) than annuals (10^{-8}–10^{-9}). De La Torre et al. (2017), analyzing transcriptomes also of different representative angiosperms, found similar rates for trees (4.0–6.0×10^{-9} silent bp substitution per year), although estimations in gymnosperms were slightly lower (generally $<1.0 \times 10^{-9}$).

These estimations, together with the high number of branches and twigs that perennial plants may produce, suggest that the number of mutations that a tree carries should be relatively high. In addition, Plomion et al. (2018), pointed out that the number of somatic mutations might be even higher since the allele frequency of these is too low to be unambiguously detected and consequently a significant proportion of mutations might not be considered. In contrast, Burian et al. (2016), using quantitative cell-lineage analysis and computational modeling, proposed that the number of fixed somatic mutations might be lower than generally calculated and that the majority of somatic mutations would be located in small sectors as nested sets of mutations. Two recent reports, in addition to Plomion et al. (2018), have provided evidence that somatic mutations certainly accumulate along branches that are developmentally connected in different sectors of the tree in a nested hierarchy analogous to a phylogenetic tree (Schmid-Siegert et al., 2017; Wang et al., 2019). These works also have clarified that mutations accumulate at a constant rate and with age, irrespective of plant stature and Xie et al. (2016), on the other hand, reported that peach (*Prunus persica* (L.) Batsch) interspecific hybrids show higher rates of somatic mutations, consistent with a former vision that increased mutation frequencies may be associated with hybridization (Emerson, 1929).

The goal of this work was to elucidate the dynamics of mosaicism in citrus, a perennial Rutaceae tree, and estimate the rates of associated mutations. The study was performed on clementine (*Citrus* ×*clementina* hort. ex Tanaka), a genetic admixture (Wu et al., 2014) whose vegetative development pattern (Garcia-Marí et al., 2002) and

physiological behavior (Iglesias et al., 2007; Tadeo et al., 2008) have been thoroughly documented because of its worldwide commercial interest.

Material and Methods

Plant Material

Samples used in this work to study somatic mutation rates were collected in July 2013 from a 36-yr-old clementine cultivar Clemenules tree, a genetic admixture of mandarin (*C. reticulata* Blanco) and pummelo (*C. maxima* (Burm.) Merr.) (Wu et al., 2014), grown in our research field at *Instituto Valenciano de Investigaciones Agrarias* in Moncada, Valencia, Spain (39°35′18.8“ N, 0°23′42.4″ W). The tree was sawn, the branches were cut, and the age of the trunk and different branches determined by dendrochronology (Supplemental Table S1). The tree, that was manually pruned during growth, showed a classic spherical crown, supported by a dominant branch of primary order (27-yr old) and two branches of secondary order (19-yr old). Citrus show a typical sympodial branch development, that in standard clementine trees of this age gives rise to approximately 2^{10} to 2^{12} branches and a number of leaves ranging from 12,000 to 34,000 (Garcia-Marí et al., 2002). This growth is mostly accompanied by a 2/5 phyllotaxis leaf arrangement. Samples consisted of all leaves from a vegetative flush (single shoot) that averaged 7.7 leaves and weighed ~3 g (Supplemental Table S2). In citrus, flushes of different ages are easily distinguishable and leaves in these flushes generally senesce and abscise in 2 yr, albeit a few of them may last one more year. Thus, somatic mutations were estimated on sprouts of three consecutive years that, for simplicity in the text, are described as flushes of 1-, 2-, and 3-yr-old. Fifteen flushes emerged either from the three main branches, from 6-yr-old branches, or from sprouts in the most basal part of the 36-yr-old trunk were harvested. In addition, two independent samples, composed of eight leaves indiscriminately picked around the tree from different orientations and heights, were also randomly collected before the experimental tree was cut to detect putative fixed mutations. As explained below, after DNA extraction, each sample was split into two technical replicates.

DNA Extraction and Whole-Genome Sequencing

In order to obtain DNA of high molecular weight and quality, an in-house nuclear isolation protocol, as described in Terol et al. (2015), was followed. In brief, fresh

leaves were homogenized using a dense buffer containing a high concentration of sucrose. The solution obtained was filtered through gauze and Miracloth layers arranged sequentially to obtain a filtrate extract that was subsequently centrifuged twice. The pellet was resuspended in a floating buffer and centrifuged again to obtain an enriched nuclei fraction on the supernatant. Nuclei recovered by pipetting were centrifuged in nuclear buffer and concentrated. The pellet was incubated overnight with RNase A and protein Kinase in a gentle stirring platform. The nuclei in the supernatant were separated from cellular debris by centrifugation and transferred to a solution with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1). A second extraction with isopropanol was carried out and DNA was recovered in the pellet, washed with ethanol, and redissolved in Tris EDTA. Finally, each sample was split into two technical replicates that were processed in parallel.

DNA libraries were constructed using standard protocols with some modifications and whole-genome sequencing was performed as essentially described in Terol et al. (2015) and Wu et al. (2018). Libraries were generated with the Illumina TruSeq DNA sample prep standard protocol and had an insert size of 500 bp. Read length was 101 bp. Paired-end sequencing was carried out on an Illumina HiSeq 2000 platform that rendered 15.7 Gb of Illumina raw data. Only raw reads containing a minimum of 70% of their nucleotides with a base quality value of at least 30 were mapped to the *C. clementina* v.1.0 haploid reference using BWA-MEM tool (Li, 2013). Map files were sorted and indexed using Samtools (Li et al., 2009). On average, 93.6% of the reference genome was covered and the mean read depth per sample was 32 (Supplemental Table S3).

Variant Calling

Variant calling was performed using the GATK-4.0.0.0 software (Van der Auwera et al., 2013). We used the HaplotypeCaller tool to generate single-sample variant call format files that were combined using the CombineGVCF tool to get a matrix including all samples sequenced in the study. Each site of the matrix showing a quality value >10 was genotyped by GenotypeGVCF and only calls tagged as SNP were filtered according to the set of standard filters specified in the variant caller practice guide. The final matrix, which included only biallelic positions, had 3,361,591 variant sites.

Somatic Mutation Identification

To retrieve only informative sites, that is, those variant sites that discriminate the genomes, the combined matrix was reduced to include relevant samples for each comparison. To avoid any bias that could be introduced in the analysis by grouping them according to their temporal references, we generated as many submatrices as pairwise comparisons could be defined. Since a sample had two technical replicates, each comparative submatrix was assembled with data coming from four different sequencing events.

The SelectVariants and VariantFiltration tools, combined with an in-house script, were used for refining the call sets. Calls had to fulfill two conditions. First, it was required that the genotype field encoding the two alleles at each site was identical between technical replicates and distinct between the two compared samples. Second, based on phred-scaled scores derived from genotype likelihoods, a genotype quality value >60 in all four targeted genomes was required.

Sites with allele balance values equal to 0 and 1 were identified as homozygous reference and homozygous variant sites, respectively. Heterozygous mutations very likely are present in all presumably identical leaves of a flush and generally produced unambiguous allele balances. For the two randomly collected samples (eight different leaves), the heterozygous threshold was established at 0.08, a value that implies that for a coverage of 32x, the alternate allele must be present in at least 2–3 reads. In addition, all putative differential calls surrounded by another one within a window of ±50 bp were ruled out because we observed that they generally are false-positive calls usually paralog sequences wrongly mapped (Li, 2011).

The analyses of technical replicates designed for mutation rate determination revealed a total of 15,809 putative calls. After visual inspection with the Integrative Genomics Viewer (IGV) browser (Robinson et al., 2011) and a careful manual curation of each one of the 15,809 calls, a final set of 67 sites was unequivocally identified. In general, the vast majority of false positives were ‘homozygous’ calls that actually were heterozygous sites, since the visual inspection revealed that at least one of the two technical replicates showed reads with an alternative allele. This usually was due to the wrong mapping or poor quality of the reads supporting the alternative allele that eventually do not fulfill the requirements of the quality filters applied and consequently

were automatically discarded by the variant caller. The analysis of the two random samplings designated to detect putative fixed mutations that were similarly processed revealed six additional mutations, thus rendering a total of 73 variants in the clementine tree (Supplemental Tables S4 and S5).

Variant Validation

The presence of the in silico detected variants was validated by polymerase chain reaction (Supplemental Table S6) and Sanger sequencing in a representative set of 12 sites. Each site was tested in samples positively identified as variants in the whole-genome sequencing data and in the two, or at least in one of the technical replicates of a sample that was homozygous for this target site. Those 12 variants were confirmed in all samples except three variants that were not validated in one sample of the random sampling analysis (Supplemental Table S7). For the validation of low allele frequencies, the consistent and recurrent occurrence of these variants in samples arranged hierarchically, and their simultaneous detection in other samples with at least an allele balance of 13%, were considered further factors of confirmation (Supplemental Table S8).

Error Analysis

We performed an error-rate estimation analysis to assess the effect that the whole process of somatic mutation identification had over the number of true calls finally retrieved (Supplemental Figure S1). The strategy designed relied on simulating somatic mutations (Keightley et al., 2015) in sites that were originally homozygous. Thus, for each pair of technical replicates, a file containing genome positions holding the reference nucleotide for the two alleles was produced. Neither depth conditions nor quality filters were applied for these positions. To ensure that the allele balance distribution of the simulation reflected the allele balance distribution of the actual data, the heterozygous positions of the whole set of genomes were first grouped in different series according to their depth values. The allele balances of the homozygous positions were then randomly assigned from the series of heterozygous positions exhibiting the same depth value. The BamSurgeon software (Ewing et al., 2015) was used to introduce these changes in the original bam files and 1,238 sites, on average, were modified in each sample. Matrices incorporating simulated positions were filtered as in the original analysis, and the number of simulated sites lost after each step was determined

(Supplemental Table S9). The average percentages of lost sites for variant calling and somatic mutation identification were 6.14 and 45.41%, respectively. To correct the number of observed mutations as related to the lost sites during somatic mutation identification, we defined a general correction factor simply dividing the number of simulated lost sites by the number of kept ones: for example, 0.4541/(1−0.4541).

Callable Sites

To define the search space of the analysis, the total number of nucleotides mapped to the reference genome was computed for each sample using Samtools (Li et al., 2009), and an averaged value of 276,277,344 nucleotides was obtained (Supplemental Table S3). This space was corrected by subtracting the average percentage of sites lost in any of the variant calling steps (5.36%; Supplemental Table S9). Therefore, the average number of callable sites was 261,457,302 nucleotides per genome that, once multiplied by the number of samples considered in the analysis, determined the search space.

Clustering Somatic Mutations

Using a presence–absence somatic mutation matrix, a phylogenetic tree was built based on the proportion of shared alleles between pairs of samples. The neighbor joining clustering method was used to draw the relationship between samples in a rooted tree using R packages (R Core Team, 2018) ggplot (Wickham, 2016), ape (Paradis and Schliep, 2019), and ggtree (Yu et al., 2017).

Mutation Rate

Mutation rate was computed dividing the number of somatic mutations detected by two times the search space.

Total Number of Mutations

To estimate the total number of variants present in clementine, the formula reported in Schoen and Schultz (2019) was applied. In this approach, the total number of mutations can be easily estimated as the product of the mutation rate, the size of the genome (Wu et al., 2014), the number of mitotic divisions that give rise to a new meristem (Burian et al., 2016), and the number of flushes or branches produced. The number of branches was estimated from the counts of total number of leaves (12,000–34,000) of a typical

clementine tree, as reported previously in Garcia-Marí et al. (2002), divided by the average number of leaves (7.7) of a single flush as calculated in Supplemental Table S2.

Results and Discussion

Sampling and Sequencing

The goal of this work was to study the dynamics of single-nucleotide mosaicism including the estimation of somatic mutations rates in citrus, an evergreen perennial fruit crop tree. We employed genome sequencing to search for somatic mutations in the clementine mandarin genome, a genetic admixture of pure mandarin and pummelo species (Wu et al., 2014). The plant material used consisted of a 36-yr-old clementine tree and somatic mutations were searched on 15 replicated samples consisting of 1-, 2-, and 3-yr-old leaf flushes. In addition, two random samples, also replicated twice, were harvested to identify putative fixed mutations. Samples were paired-end sequenced with an Illumina HiSeq 2000 platform at an average coverage of 32x (Supplemental Table S3).

Identification of Somatic Mutations

The experimental design used for the detection of somatic mutations involved a series of sequential steps including variant calling, somatic mutation identification, IGV browser confirmation, and Sanger validation (Supplemental Table S7) as described in the Material and Methods section (Supplemental Figure S1). For a coverage of 32x, variants present in a flush (presumably eight identical leaves) generally had an unambiguous allele balance between 0.15 and 0.44 with a few determinations above and below this range (Supplemental Table S8). Allele balance is not close to 0.5, presumably because mutations may occur in specific cell layers. For the alternate allele in the two randomly collected samples (eight leaves from different flushes), allele balance ranged between 0.1 and 0.2, again with a few determinations above and below these values. Validation of low frequency variants required, in addition to the general strategy followed in the identification process, the consistent and recurrent occurrence of these variants in samples arranged hierarchically and their simultaneous detection in other samples with at least an allele balance of 13% (Supplemental Table S8). The error estimation of the whole mutation detection process was calculated through mutation simulation and subsequent computational analyses as performed with the real variants

(Supplemental Table S9). Two correction factors, considering the search space and the percentage of variants lost in every step of the bioinformatics pipeline, were finally applied.

The analysis of 15 pairs of technical replicates designed for mutation rate determination rendered 105 submatrices including a total of 15,089 calls. All these calls were visually inspected using the IGV browser, and after a manual careful curation of each one of these 15,089 calls, a final set of 67 sites was unequivocally identified. The analysis of the two random samplings designated to detect putative fixed mutations, that were similarly processed, revealed six additional mutations, thus rendering a total of 73 variants (Supplemental Tables S4 and S5).

Distribution and Types of Nucleotide Substitutions

In the 17 samples analyzed, we identified 73 reliable somatic mutations, a higher part of which, 39, were shared by at least two different samples while the remaining 34 base substitutions were singleton SNPs. The clementine genome annotation (https://phytozome.jgi.doe.gov/pz/portal.html#!bulk?org=Org_Cclementina) was used to categorize the detected sites into functional classes. As expected, many mutated sites were found in intergenic (53) and intronic (11) regions and therefore had no effect on protein sequence, while nine substitutions were registered in coding regions and five of them were nonsynonymous (Supplemental Table S5). These frequencies are in line with previous reports supporting the idea that somatic mutations are not affected by selection (Ossowski et al., 2010; Wang et al., 2019).

The distribution of the somatic mutations was rather uniform among the nine clementine chromosomes (Figure 1). Within chromosomes, SNPs were found evenly distributed between mandarin–mandarin regions and pummelo introgressed fragments and also between areas of high and low heterozygosity. Similarly, the data show that the low genic regions (centromere, pericentromeric, and transposon areas; Borredá et al. 2019) contain the same number of nucleotide substitutions (49%) that are found in genic regions. Overall, these data indicate that the generation of somatic mutations is a stochastic process.

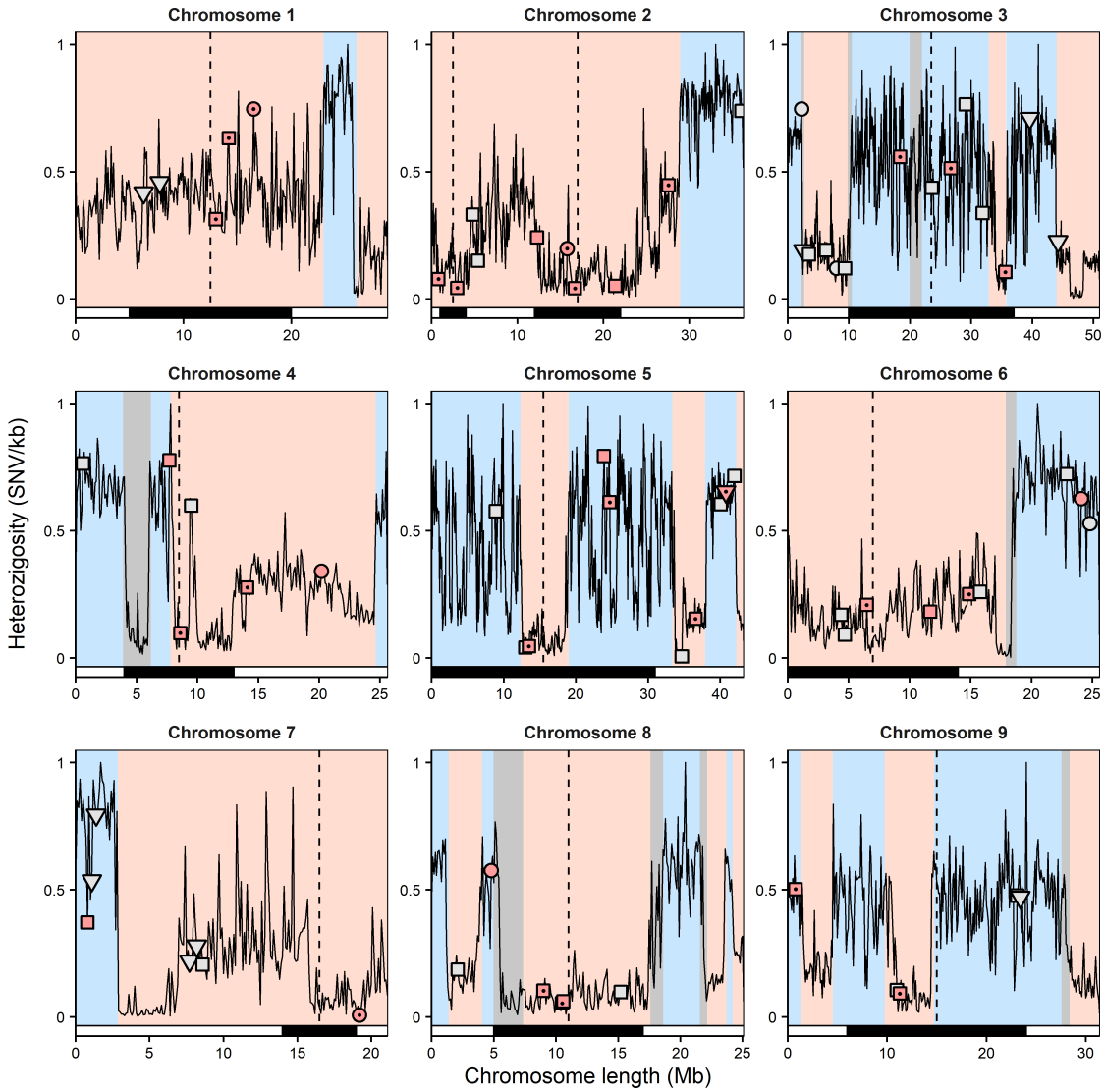


Figure 1. Distribution of somatic mutations identified in a 36-yr-old clementine tree along the chromosomes. Heterozygosity is represented in the y axis as the nucleotide diversity relative frequency. Pink areas correspond to mandarin–mandarin stretches; blue areas correspond to introgressed regions heterozygous for mandarin–pummelo genotype; grey areas, in principle are unknown areas. Black blocks in the horizontal bars represent low genic regions, including centromeres, pericentromeric, and transposon regions, while white blocks represent high genic regions (Borredá et al., 2019). Symbols code: circle, coding sequence (CDS); inverted triangle, introns; square, intergenic. Color code: red, G/C to A/T transitions; grey, any nucleotide substitution different from G/C to A/T transitions; point inside the symbol, G/C to A/T transitions in dipyirimidine sites.

The analyses of the type of nucleotide substitution observed indicate that out of the 73 SNPs detected, 52 were transitions and 21 transversions (Figure 2; Supplemental Table S5), and a high proportion of them, 35 (48%), were G/C to A/T transitions, a frequency in the range reported in other plants (Ossowski et al., 2010). The detection of this

specific type of substitutions suggests exposure to UV that produces both base substitutions of cytosine to thymine at dipyrimidine sites and deamination of methylate cytosines at CpG sites that also renders thymines (Friedberg et al., 2006). The first pattern of mutations conforms to the so-called UV signature, while the second one is mainly restricted to the solar UV signature (Ikehata and Ono, 2011). In our set of data, 25 SNPs of the 35 G/C to A/T transitions were detected in a dipyrimidine context (Supplemental Tables S5 and S10.

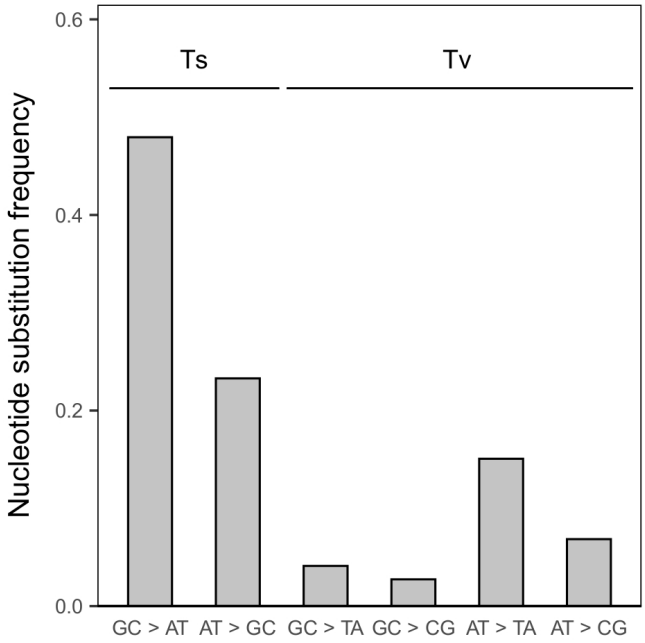


Figure 2. Frequency of nucleotide substitution type identified in flushes emerged from the main trunk and branches of different ages in a 36-yr-old clementine tree. Ts, transition; Tv, transversion. Arrow (>) indicates the direction of the nucleotide substitution.

Dynamics of Single-Nucleotide Mosaicism

In order to detect somatic mutations and to study their occurrence, distribution, and persistence in different locations along the clementine tree, we harvested 15 samples consisting of 1, 2, and 3-yr-old flushes sprouted in branches of different ages. A rooted neighbor-joining tree based on shared allele distance between the spontaneous mutations identified in the samples revealed four main clades (I, II, III and IV) (Figure 3), basically corresponding to the four scenarios that were originally sampled. The tree presented two superior clusters: one included Clades I and IV and the other included Clades II and III. A fourth basal leaf flush, with no connections with any other flushes,

was also isolated and hence used as an outgroup of the tree. Clade I nested SNPs on two leaf flushes emerged in the basal part of the 36-yr-old trunk of the tree. Clade II grouped mutations detected on the 27-yr-old branch. The remaining two clades, Clades III and IV, included mutations grouped on each one of the other 19-yr-old secondary order branches, both carrying also a 6-yr-old branch. Clade IV was linked to Clade I, while Clade II was nested first with one of the flushes emerged from the basal part of the trunk and secondly with Clade III.

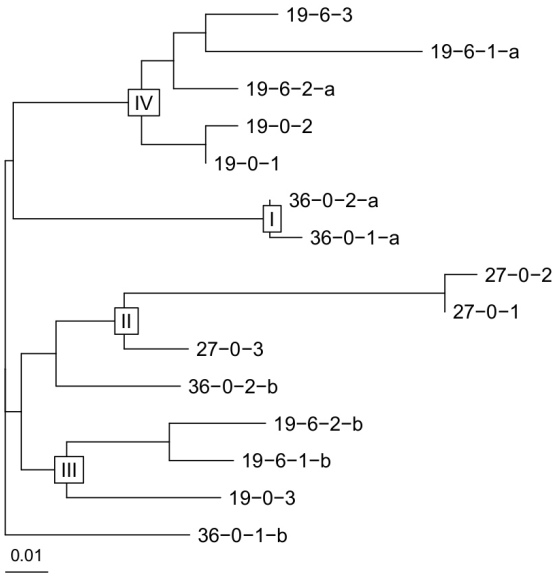


Figure 3. Rooted neighbor-joining tree based on the proportion of shared alleles between pairs of samples. Somatic mutations were detected in leaf flushes emerged from the main trunk and branches of different ages in a 36-yr-old clementine tree. Roman numerals specify the main four clades revealed by the cluster analysis. Sample code: first number, age of the primary or secondary order branch; second number, age of the following order branch; third number, age of the flush; the final letter, if needed, discriminates samples with identical sequence of branching events.

Figure 4 shows a schematic representation of our experimental tree, locating the 67 mutations identified in the flush analyses. The distribution of the same mutations in sectorized sections of the tree, that is, on several branches and leaf flushes physically connected, indicates that somatic mutations that originally arose in an initial cellular division of the meristem were retained and transmitted during the growth of the different structures. There were SNPs, for instance, detected in the basal parts of the tree that were recurrently found in different branches and leaf flushes irrespectively of their ages. These somatic mutations that spread throughout the new growth, colonizing all

organs along a particular sector, can in fact be considered fixed somatic mutations of this sector (see below). The relationship between Clades II and III constitute an illustrative example of this idea. The three leaf flushes (1, 2, and 3-yr-old) emerged on the 27-yr-old branch in Clade II contained at least two SNPs (dark green) that were identified in a 2-yr-old flush sprouted in the basal part of the 36-yr-old trunk of the tree. Additionally, this flush carried two SNPs (orange and cream) detected in the three samples analyzed of the 19-yr-old branch in Clade III. The older flush in the main branch of this clade contained an additional SNP (red) also identified in the basal flush. These results indicate that the somatic mutations detected in the 2-yr-old flush of the basal part of the tree were fixed, two of them (dark green) in the branch of 27 yr and two others (orange and cream) in the flushes sprouted on the 19-yr-old branch. Furthermore, this basal flush included a variant (cream) that was found in all sprouts of Clades I, III, and IV, but not on Clade II. The observation that a particular SNP emerged earlier in a 3-yr-old flush in the upper part of the tree than in a 2-yr-old flush generated in the lower part of the trunk reveals that this somatic mutation was generated during the first divisions that gave rise to this trunk and was ‘dragged’ during the growth of the trunk to form older branches. This mutation, on the other hand, was also kept in a dormant bud that sprouted forming the 2-yr-old flush emerged in the basal trunk.

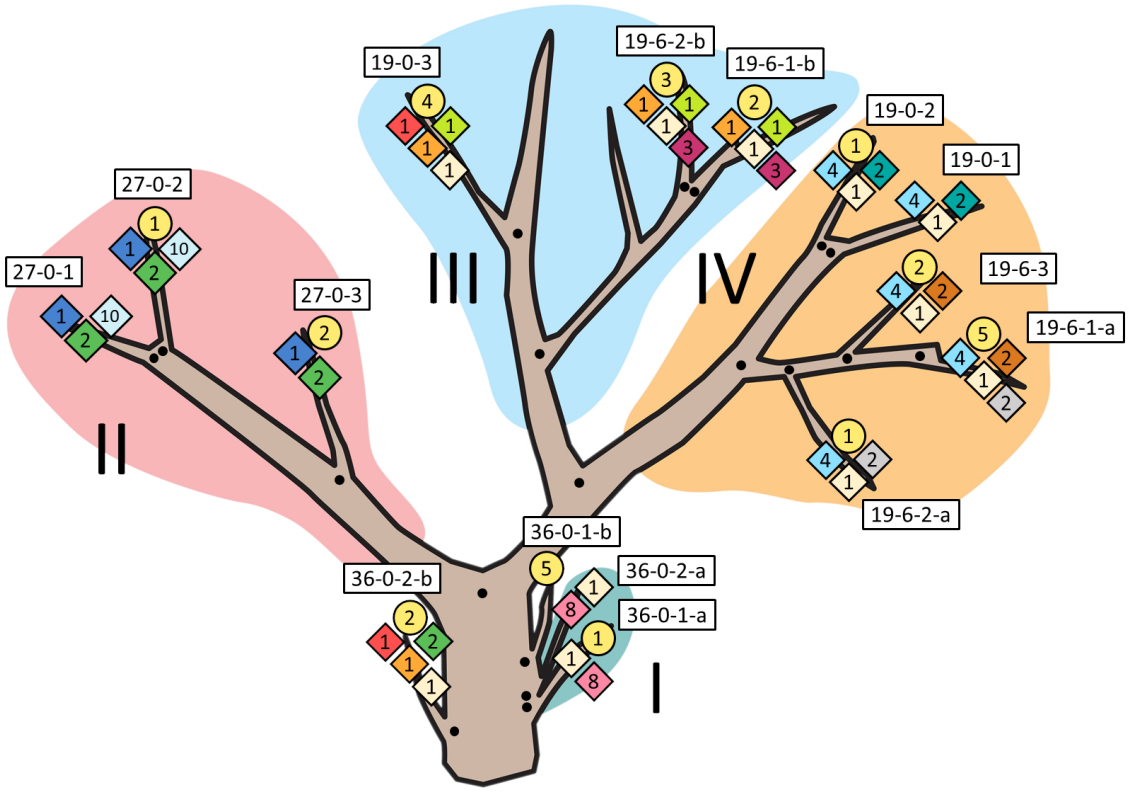


Figure 4. Schematic representation of the 36-yr-old clementine tree used for sequencing to study the dynamics of single-nucleotide mosaicism in citrus. Somatic mutations were searched on samples consisting of 1-, 2- and 3-yr-old leaf flushes emerged from 27-, 19- and 6-yr-old branches and from the most basal part of the 36-yr-old trunk. Roman numerals indicate the four scenarios revealed by the neighbor-joining tree based on the proportion of shared alleles between pairs of samples (Figure 3). Numbers in rectangles correspond to the codes of the samples sequenced. Sample code: first number, age of the primary or secondary order branch; second number, age of the following order branch; third number, age of the flush; the final letter, if needed, discriminates samples with identical sequence of branching events. Numbers in yellow circles indicate the number of singleton single-nucleotide polymorphisms (SNPs) detected in each flush. Number in squares indicate the number of shared SNPs detected in different flushes and colors identify the flushes that share the same SNPs. Black dots indicate branching events.

In Clade IV, supported by a branch of 19 yr and a secondary branch of 6 yr, there were at least five variants (four blue and one cream) that were found in all five flushes analyzed and therefore were fixed mutations in this sector (see below). Other SNPs that were shared only between flushes emerged in the same branch (either 19- or 6-yr-old), probably arose in the buds that give rise to these flushes or branches. Other different scenarios in Figure 4 can be described that, in essence, exemplify the same dynamic of single-nucleotide mosaicism, confirming that somatic mutations are arranged in a nested hierarchy (Schmid-Siegert et al., 2017). The data may indicate that somatic mutations can be fixed in superior order branches (see below), and so on in an iterative process, a

situation predicted by computational analyses in Burian et al. (2016) that suggested that the bulk of variants are distributed in small sectors.

Estimations of Mutation Rates

Albeit mutation rates are generally calculated in a per-time basis, that is, per year or per generation, somatic mutation accumulation, as exemplified above, relies on a progressive iterative process that deserves further analysis. This vision suggests that the expected number of somatic mutations is critically dependent upon the number of accumulated cell divisions produced during the branching process, that is, on the number of bifurcations that eventually bring about a certain branch (Schoen and Schultz, 2019). The estimate of the rate of somatic mutations in our experimental tree, for instance, was 4.4×10^{-10} bp^{-1} yr^{-1} (Table 1), a value practically identical ($4.0–6.0 \times 10^{-10}$) to that reported in other fruit crops of similar ages, such as peach (Wang et al., 2019), and slightly higher than those found in Sitka spruce (5.3×10^{-11} to 1.2×10^{-10}), a centenary conifer, more genetically distant (Hanlon et al., 2019). Our estimate is a corrected value, calculated following the procedure detailed in the methods section, based on the 67 nucleotide substitutions found in the 15 leaf flushes (1, 2, and 3 yr) sequenced. Further insights on the accumulated mutation rates in branches of different ages (36, 27, 19, and 6 yr), including all SNPs detected in the sprouts of each branch, indicated that these values were higher as the branch age was lower. Thus, a negative correlation ($y = 4.0 \times 10^{-9}$, $e^{-0.061x}$, $R^2 = 0.93$, $p = 0.04$) was obtained between the branch age and the rate of the accumulated mutations in these branches (Figure 5). These data provide evidence that somatic mutations accumulate at constant rates during time in the different branches and therefore that the buildup of mutations should be rather uniform among branches of the same ages and number of branching bifurcations. These results are not unexpected, since younger branches in a sympodial pattern of branching sprout upon older branches, thus increasing the potential ‘to drag’ and therefore to accumulate somatic mutations in the flushes emerged in younger branches. The iterative sympodial process that enhances the number of branching points increases, therefore, the number of cell divisions and the number of accumulated mutations. As a result, mutations tend to accumulate with age, suggesting that a very huge number of mutations can be expected to accumulate in the younger branches of the tree.

Table 1. Accumulated somatic mutation rates estimated in a 36-yr-old clementine tree. Note: The total number of variants detected in all (1-, 2- and 3-yr-old) leaf flushes analyzed was used for the average estimation of nucleotide substitutions of the whole tree. Variants found in flushes emerged from the most basal part of the trunk tree (sprouts), and from 27-, 19- and 6-yr-old branches were used for the average estimation of accumulated nucleotide substitutions in trunk sprouts and in older, middle and younger branches. The search space, the raw rates, and the calculated rates after corrections specified in the Material and Methods section, are also shown.

Samples	**Estimated age (years)**	**Number of nucleotide substitutions**	**Search space (sites)**	**Accumulated mutation rates (Nucleotide substitutions)**		
				Raw	**Corrected**	**Per year**
Whole tree	36	67	3,921,859,535	8.54×10^{-9}	1.56×10^{-8}	4.35×10^{-10}
Trunk sprouts	36	21	1,045,829,209	1.00×10^{-8}	1.84×10^{-8}	5.11×10^{-10}
Older branches	27	16	784,371,907	1.02×10^{-8}	1.87×10^{-8}	6.92×10^{-10}
Middle branches	19	15	784,371,907	9.56×10^{-9}	1.75×10^{-8}	9.22×10^{-10}
Younger branches	6	27	1,307,286,512	1.03×10^{-8}	1.89×10^{-8}	3.15×10^{-9}

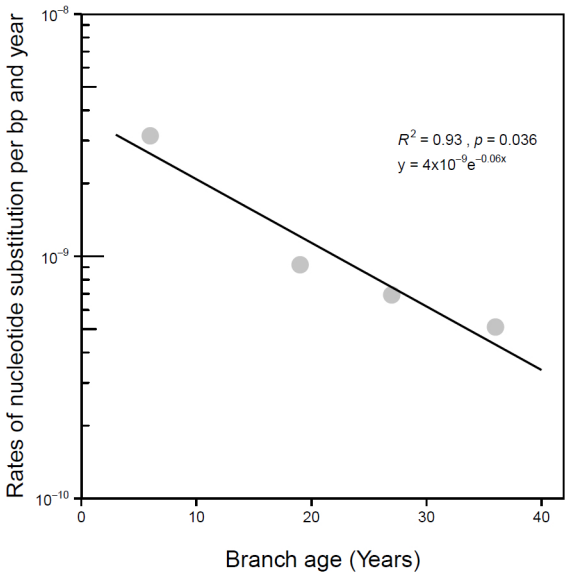


Figure 5. Accumulated mutation rate correlation with the age of the branches. Mutation rates were estimated in leaf flushes emerged from 27-, 19-, and 6-yr-old branches and from the most basal part of the trunk of a 36-yr-old clementine tree. All different variants detected in the flushes of each branch were considered for the analysis. Exponential regression provided a better fit than simple linear regression.

Fixed Mutations

In principle, the presence of somatic mutations fixed in all branches of the tree should be very rare since the pattern of variant accumulation described above determines that the vast majority of variants should be present in separate sectors along the tree (Figure 4). Consistently, there were no variants shared between the three branches of the three upper sectors, in our experimental tree, in line with the idea that branches derived from specific areas of the initial shoot–apical meristem (Wang et.al, 2019).

In order to identify fixed somatic mutations, that is, variants present in entire sectors of the tree, two independent random samplings consisted each one of eight nonrelated leaves, harvested from different orientations and heights, were performed before the experimental tree was cut. In these analyses, 12 mutations out the 18 variants detected in these random samplings were previously identified in the Scenarios II (27-yr-old branch; seven variants), III (19-yr-old branch; one variant), and IV (19-yr-old branch; two variants), while a single variant was found in three scenarios (I, III, and IV) and another one in two scenarios (I and II) (Figure 4). It is explanatory to mention that the 27-yr-old primary branch was the oldest and dominant branch of the tree, an observation related to the higher association of the variants coming from the random samplings with

Clade II. The allele balance of seven of these somatic mutations, present in Scenarios II, or IV, or II, III, and IV was roughly 0.4, indicating a prevalence of the variants in almost all leaf cells of the flush sequenced (Supplemental Table S8). In the random samplings, this value generally ranged between 0.1 and 0.2, since the sample included leaves from different sectors of the tree, which diluted the balance. Furthermore, three of this set of seven somatic mutations were also detected in two sprouts of the trunk, supporting their extensive dispersion in the tree. These observations suggest that all these seven mutations are widespread within the tree and can therefore be considered real fixed mutations, although their presence is restricted to specific sectors, that is, they have not been propagated to the whole tree. The six singleton variants identified in these analyses are probably not fixed mutations since they were present at low frequencies; five of them were found only in one of the two random samples and were not detected in the determinations of mutations rates. Obviously, there must be many other mutations that escaped our survey, as we only covered samples that represent at most 31 leaf shoots out of the thousands that a tree may carry.

Number of Total Mutations

Therefore, another question of relevance is related to estimation of the number of total somatic mutations in a tree. For a long time, it has been believed that the number of somatic mutations in citrus should be relatively high since the vast majority of the thousands of known varieties, clones, or cultivars of citrus have invariably been selected from spontaneous mutations.

Schoen and Schultz (2019), in a recent review have pointed out that genomic determinations of mutation rates in plants are at least two orders of magnitude lower than those estimated via the currently accepted rates of cell division and mitotic mutation in model organisms. Plomion et al. (2018), for instance, argue that most somatic mutations are expected to remain at frequencies too low to be unambiguously detected and therefore that it still remains particularly challenging to determine the actual rate of somatic mutations. This view suggests that mutation rates based on sequencing approaches tend to underestimate the authentic number of variants present in a tree.

To provide approximations on the total number of variants existing in clementine trees, we have used previous estimates on the number of leaves present in representative

clementine crowns and applied the formula reported in Schoen and Schultz (2019). According to this report, the expected number of cumulative base-pair mutations is the product of the genome size (in base pairs), the number of base-pair mutations per stem cell division, the number of stem cell divisions required to produce an axillary meristem that subsequently gives rise to a flush and to a branch and the number of branches produced. In our calculations, the number of branches was obtained dividing the average number of leaves of a typical clementine tree, ranging from 12,000 to 34,000 (Garcia-Marí et al., 2002), by the average number of leaves of a single flush (7.7; Supplemental Table S2). These estimates render a quantity of branches, between 1,558 and 4,416, that is in the range of the 2^{10}–2^{12} branching bifurcations (1,024–4,096 branches) that usually develops a standard clementine tree of similar age. Therefore, if we consider a perfect, ideal configuration tree, applying the mutation rate found in this work, 4.4×10^{-10} bp^{-1} yr^{-1} (Table 1), assuming 3×10^{8} bp in the clementine genome (Wu et al., 2014), that seven to nine cell divisions separate the branching events independent of the branch age (Burian et al., 2016) and that the number of branches of a typical clementine tree may vary between 1,558 and 4,416, the expected number of base-pair mutations present in the crown would rank from a minimum of 1,440 (seven cells and 12,000 leaves) to a maximum of 5,246 mutations (nine cells and 34,000 leaves). From these data, the average number of mutations of an axillary meristem or branching bifurcation is very close to 1 (0.92–1.19) and since a bifurcation generates a flush or a branch, this also is the estimated number of total variants per flush and even per leaf, assuming that all leaves in a flush are genetically identical. This relatively high number of mutations may be associated with the enormous number of citrus varieties of spontaneous origin that have been described worldwide. It is worth mentioning that in addition to the novel phenotypes generated by sport mutations, the selection of citrus nucellar seedlings with superior performance, an old practice in citriculture, may also be related to positive and negative selection of the suits of somatic mutations that are expressed in trees derived from the nucellar embryos.

In the perennial citrus, leaves are generally renewed every 2 yr, a circumstance that may have evolutionary consequences because it allows renovation and substitutions of mutations not previously fixed in the tree. The absence of sexual reproduction eventually may lead to processes of mutational meltdowns or Muller's ratchets resulting in the accumulation of deleterious and harmful mutations in an irreversible manner. The

sectorial distribution of somatic mutations and the periodic leaf substitution in the long term may increase genetic heterogeneity and therefore the adaptive role of somatic mutations (Whitham and Slobodchikoff, 1981) reducing, in part, the chances of mutational load. Intraorganismal heterogeneity is required, for instance, by the genetic mosaicism hypothesis (Gill et al., 1995) that proposes that spontaneous mutations provide direct fitness benefits (Folse and Roughgarden, 2012), conferring insect and herbicide resistance (Simberloff and Leppane, 2019), an essential aspect of plant–herbivore interactions and hence, of the coevolution between long-lived trees and short-lived herbivores.

In conclusion, we provide evidence that somatic mutations in citrus, a long-live perennial, spread following an iterative pattern determined by the sympodial model of branching and are consequently found grouped in sectors of the tree where some of them become fixed. Many of the 73 reliable identified mutations were G/C to A/T transitions, a type of nucleotide substitution that appears to be associated with UV exposure. The average estimate of mutation rates in our experimental tree was 4.4×10^{-10} bp^{-1} yr^{-1} and the data also indicated that the rates of accumulated mutation in branches were higher as the branch age was lower. Since the number of leaves or flushes are a function of the number of branching bifurcations, the total number of mutations on a tree can be estimated if the mutation rates and the number of leaves, flushes, or branching points are known. Thus, assuming a perfect, ideal configuration, a clementine tree should carry a total of 1,500–5,000 variants, while each axillary meristem appears to produce one somatic mutation, on average. These relatively elevated number of mutations may also be in line with the huge number of varieties derived from spontaneous mutations that are commercialized in citrus. From an evolutionary standpoint, the sectorial distribution of mutations and the periodic leaf renewal habit of citrus appear to increase genetic heterogeneity, and therefore, somatic mutations may play an adaptive role in reducing mutational load.

Acknowledgements

This research is co-funded by the Ministerio de Ciencia, Innovación y Universidades (Spain) trough grant # RTI2018-097790-R-100 and by the European Union through the European Regional Development Fund (ERDF) of the Generalitat Valenciana 2014-2020, through grants IVIA 51915 and 52002. We thank Matilde Sancho, Isabel Sanchis and Antonio Prieto for laboratory tasks.

Author Contribution

MT: Conceptualization, Funding acquisition, Investigation, Supervision, Writing-original draft, Writing-review & editing; EPR: Data Curation, Investigation, Formal analysis, Visualization; Writing-original draft; ALGU: Resources; CB: Data curation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Sequence data used in the analysis are publicly available at the NCBI Sequence Read Archive, under BioProject PRJNA762220.

References

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., Levy, Y., Harel, T. H., Shalev-Schlosser, G., Amsellem, Z., Razifard, H., Caicedo, A. L., Tieman, D. M., Klee, H., Kirsche, M., Aganezov, S., … Lippman, Z. B. (2020). Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell, 182*(1), 145–161.e23. https://doi.org/10.1016/j.cell.2020.05.021

Borredá, C., Pérez-Román, E., Ibanez, V., Terol, J., & Talon, M. (2019). Reprogramming of Retrotransposon Activity during Speciation of the Genus Citrus. *Genome biology and evolution, 11*(12), 3478–3495. https://doi.org/10.1093/gbe/evz246

Burian, A., Barbier de Reuille, P., & Kuhlemeier, C. (2016). Patterns of Stem Cell Divisions Contribute to Plant Longevity. *Current biology*, 26(11), 1385–1394. https://doi.org/10.1016/j.cub.2016.03.067

De La Torre, A. R., Li, Z., Van de Peer, Y., & Ingvarsson, P. K. (2017). Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. *Molecular biology and evolution,* 34(6), 1363–1377. https://doi.org/10.1093/molbev/msx069

Emerson R. A. (1929). The Frequency of Somatic Mutation in Variegated Pericarp of Maize. *Genetics, 14*(5), 488–511.

Ewing, A. D., Houlahan, K. E., Hu, Y., Ellrott, K., Caloian, C., Yamaguchi, T. N., Bare, J. C., P'ng, C., Waggott, D., Sabelnykova, V. Y., ICGC-TCGA DREAM Somatic Mutation Calling Challenge participants, Kellen, M. R., Norman, T. C., Haussler, D., Friend, S. H., Stolovitzky, G., Margolin, A. A., Stuart, J. M., & Boutros, P. C. (2015). Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nature methods*, *12*(7), 623–630. https://doi.org/10.1038/nmeth.3407

Folse, H. J., 3rd, & Roughgarden, J. (2012). Direct benefits of genetic mosaicism and intraorganismal selection: modeling coevolution between a long-lived tree and a short-lived herbivore. *Evolution; international journal of organic evolution, 66*(4), 1091–1113. https://doi.org/10.1111/j.1558-5646.2011.01500.x

Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A. & Ellenberger, T. (2006). *DNA Repair and Mutagenesis* (Second Edition). ASM (American Society for Microbiology) Press. Washington, DC.

Frost, H. B., & Krug, C. A. (1942). Diploid-Tetraploid Periclinal Chimeras as Bud Variants in Citrus. *Genetics*, 27(6), 619–634.

Garcia-Marí, F., Granda, C., Zaragoza, S., & Agustí, M. (2002). Impact of Phyllocnistis citrella (Lepidoptera: Gracillariidae) on leaf area development and yield of mature citrus trees in the Mediterranean area. *Journal of economic entomology, 95*(5), 966–974. https://doi.org/10.1093/jee/95.5.966

Gill, D. E., Chao, L., Perkins, S. L. & Wolf, J. B. (1995). Genetic mosaicism in plants and clonal animals. Annual Review of Ecology, Evolution, and Systematics. 26, 423–444. https://doi.org/10.1146/annurev.es.26.110195.002231

Hanlon, V., Otto, S. P., & Aitken, S. N. (2019). Somatic mutations substantially increase the per-generation mutation rate in the conifer Picea sitchensis. *Evolution letters, 3*(4), 348–358. https://doi.org/10.1002/evl3.121

Iglesias, D. J., Cercós, M., Colmenero-Flores, J. M., Naranjo, M. A., Ríos, G., Carrera, E., et al., (2007). Physiology of citrus fruiting. Braz. *J. Plant Physiol. 19*(4), 333-362.

Ikehata, H., & Ono, T. (2011). The mechanisms of UV mutagenesis. *Journal of radiation research, 52*(2), 115–125. https://doi.org/10.1269/jrr.10175

Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., Davey, J. W., & Jiggins, C. D. (2015). Estimation of the spontaneous mutation rate in Heliconius melpomene. *Molecular biology and evolution, 32*(1), 239–243. https://doi.org/10.1093/molbev/msu302

Klekowski, E., & Godfrey, P. (1989). Ageing and mutation in plants. *Nature 340*, 389–391. https://doi.org/10.1038/340389a0

Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England), 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lupski J. R. (2013). Genetics. Genome mosaicism--one human, multiple genomes. *Science (New York, N.Y.), 341*(6144), 358–359. https://doi.org/10.1126/science.1239503

National Library of Medicine (US), National Center for Biotechnology Information; (2004). Available at: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Citrus_clementina/100

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D., & Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science (New York, N.Y.), 327*(5961), 92–94. https://doi.org/10.1126/science.1180677

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics (Oxford, England), 35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Plomion, C., Aury, J. M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillonne, N., Labadie, K., Le Provost, G., Lesur, I., Bartholomé, J., Faivre-Rampant, P., Kohler, A., Leplé, J. C., Chantret, N., Chen, J., Diévart, A., Alaeitabar, T., Barbe, V., Salse, J. (2018). Oak genome reveals facets of long

lifespan. *Nature plants, 4*(7), 440–452. https://doi.org/10.1038/s41477-018-0172-3

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology, 29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Schmid-Siegert, E., Sarkar, N., Iseli, C., Calderon, S., Gouhier-Darimont, C., Chrast, J., Cattaneo, P., Schütz, F., Farinelli, L., Pagni, M., Schneider, M., Voumard, J., Jaboyedoff, M., Fankhauser, C., Hardtke, C. S., Keller, L., Pannell, J. R., Reymond, A., Robinson-Rechavi, M., Xenarios, I., … Reymond, P. (2017). Low number of fixed somatic mutations in a long-lived oak tree. *Nature plants, 3*(12), 926–929. https://doi.org/10.1038/s41477-017-0066-9

Schoen, D. J. & Schultz, S. T. (2019). Somatic mutation and evolution in plants. *Annual Review of Ecology, Evolution, and Systematics, 50*(1), 49-73

Simberloff, D., & Leppanen, C. (2019). Plant somatic mutations in nature conferring insect and herbicide resistance. *Pest management science, 75*(1), 14–17. https://doi.org/10.1002/ps.5157

Szymkowiak, E. J., & Sussex, I. M. (1996). What chimeras can tell us about plant development *Annual review of plant physiology and plant molecular biology, 47*, 351–376. https://doi.org/10.1146/annurev.arplant.47.1.351

Tadeo, F.R., Cercós, M., Colmenero-Flores, J. M., Iglesias, D. J., Naranjo, M. A., Ríos, G., Carrera, E., Ruiz-Rivero, O., Lliso, I., Morillon, R., Ollitrault, P., Talon, M. (2008). Molecular Physiology of Development and Quality of Citrus. *Advances in Botanical Research*. *47*, 147–223. https://doi.org/10.1016/S0065-2296(08)00004-9

Terol, J., Ibañez, V., Carbonell, J., Alonso, R., Estornell, L. H., Licciardello, C., Gut, I. G., Dopazo, J., & Talon, M. (2015). Involvement of a citrus meiotic recombination TTC-repeat motif in the formation of gross deletions generated by ionizing radiation and MULE activation. *BMC genomics, 16*(1), 69. https://doi.org/10.1186/s12864-015-1280-3

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics, 43*(1110), 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Wang, J., Li, X., Do Kim, K., Scanlon, M. J., Jackson, S. A., Springer, N. M., & Yu, J. (2019). Genome-wide nucleotide patterns and potential mechanisms of genome divergence following domestication in maize and soybean. *Genome biology, 20*(1), 74. https://doi.org/10.1186/s13059-019-1683-6

Wang, L., Ji, Y., Hu, Y., Hu, H., Jia, X., Jiang, M., Zhang, X., Zhao, L., Zhang, Y., Jia, Y., Qin, C., Yu, L., Huang, J., Yang, S., Hurst, L. D., & Tian, D. (2019). The

architecture of intra-organism mutation rate variation in plants. *PLoS biology, 17*(4), e3000191. https://doi.org/10.1371/journal.pbio.3000191

Whitham, T. G., & Slobodchikoff, C. N. (1981). Evolution by individuals, plant-herbivore interactions, and mosaics of genetic variability: The adaptive significance of somatic mutations in plants. *Oecologia, 49*(3), 287–292. https://doi.org/10.1007/BF00347587

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York

Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences of the United States of America, 84(24), 9054–9058. https://doi.org/10.1073/pnas.84.24.9054

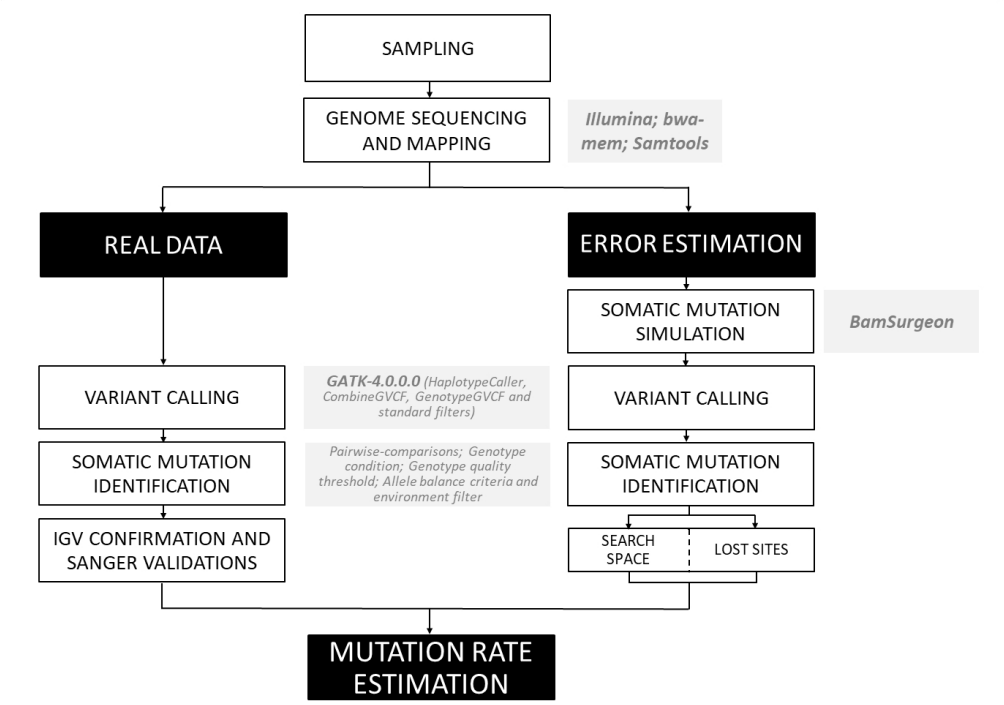
Wu, G. A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., Takita, M. A., Labadie, K., Poulain, J., Couloux, A., Jabbari, K., Cattonaro, F., Del Fabbro, C., Pinosio, S., Zuccolo, A., Chapman, J., … Rokhsar, D. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature biotechnology, 32*(7), 656–662. https://doi.org/10.1038/nbt.2906

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., Curk, F., Du, D., Ollitrault, P., Roose, M. L., Dopazo, J., Gmitter, F. G., Rokhsar, D. S., & Talon, M. (2018). Genomics of the origin and evolution of Citrus. *Nature, 554*(7692), 311–316. https://doi.org/10.1038/nature25447

Xie, Z., Wang, L., Wang, L., Wang, Z., Lu, Z., Tian, D., Yang, S., & Hurst, L. D. (2016). Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. A low rate in woody perennials and a higher mutagenicity in hybrids. *Proceedings. Biological sciences, 283*(1841), 20161016. https://doi.org/10.1098/rspb.2016.1016

Yu, G., Smith, D., Zhu, H., Guan, Y. & Tsan-Yuk Lam, T (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28-36.

Supplementary Material



Supplemental Figure S1. Experimental design used for the detection of somatic mutations in a 36-yr-old tree. For the identification of variants, main steps of the approach involved variant calling, somatic mutation identification, IGV confirmation and Sanger validation. For error estimation, mutation simulations were generated and computationally processed as done for real variants. Two correction factors, that take into account the search space and the percentage of variants lost in every step of the bioinformatics pipeline, were finally applied.

Table S1. Dendrochronological analysis.

Samples	Side	Ring counting			Average	Estimated years
		X	Y	Z		
Trunk	Upper	43	38	36	39	36.5
	Lower	35	36	31	34	
Older branch	Upper	25	27	24	25.3	26.7
	Lower	31	24	29	28	
Middle branches	Upper	18	21	17	18.7	19
	Lower	19	20	19	19.3	
Younger branches	Upper	7	6	6	6.3	6.3
	Lower	6	6	7	6.3	

Table S2. Number of leaves per flush, and leaf weight.

Sample	Number of flushes	Number of leaves	Leaves per flush	Standard deviation	Leaf fresh weight (g)	Standard deviation
sample_1	9	79	8.78	-	0.40	0.16
sample_2	9	63	7	-	0.30	0.17
sample_3	8	56	7	-	0.48	0.26
sample_4	9	72	8	-	0.29	0.12
		Average	7.71	0.86	0.36	0.09

Table S3. Sample codes and whole genome sequencing statistics.

Sample codes					Number of raw reads	Number of sequenced nucleotides	Coverage	Number of mapped nucleotides	Percentage of reference genome with depth greater than 10x
1st order branch	2nd order branch	Leaf flush		Tech. Replicate					
19	6	1	a	1	146,115,248	14,757,640,048	30	276,169,834	86
				2	153,213,508	15,474,564,308	32	276,307,151	86
19	6	1	b	3	159,539,696	16,113,509,296	32	276,655,829	86
				4	136,398,054	13,776,203,454	28	276,075,360	84
19	0	1		5	125,737,912	12,699,529,112	26	275,848,611	83
				6	185,371,286	18,722,499,886	38	276,974,176	87
27	0	1		7	155,170,326	15,672,202,926	32	276,143,918	86
				8	127,651,428	12,892,794,228	26	275,859,523	83
36	0	1	b	9	148,648,792	15,013,527,992	31	276,075,414	86
				10	149,977,736	15,147,751,336	31	276,486,548	86
36	0	1	a	11	138,542,524	13,992,794,924	29	276,039,299	85
				12	136,405,098	13,776,914,898	28	275,999,022	85

19	6	2	a	13	130,332,378	13,163,570,178	26	275,676,540	83
				14	167,693,872	16,937,081,072	33	276,367,361	86
19	6	2	b	15	138,306,704	13,968,977,104	27	276,133,377	84
				16	153,857,790	15,539,636,790	31	276,514,591	85
27	0	2		17	147,989,648	14,946,954,448	30	276,002,722	85
				18	130,623,346	13,192,957,946	26	276,038,176	84
19	0	2		19	149,206,332	15,069,839,532	30	275,809,119	85
				20	142,171,646	14,359,336,246	29	275,881,582	85
36	0	2	a	21	144,494,638	14,593,958,438	30	276,071,136	85
				22	144,283,162	14,572,599,362	30	275,873,267	85
36	0	2	b	23	128,712,914	13,000,004,314	26	275,758,566	84
				24	146,413,268	14,787,740,068	30	275,944,712	85
19	6	3		25	165,830,168	16,748,846,968	33	276,615,994	86
				26	126,872,118	12,814,083,918	26	275,286,003	83
27	0	3		27	166,353,914	16,801,745,314	34	276,752,764	87
				28	143,205,808	14,463,786,608	29	276,226,898	85
19	0	3		29	160,114,340	16,171,548,340	33	276,786,708	87
				30	310,442,980	31,354,740,980	64	278,561,026	90

Random sampling	a	31	131,160,314	13,247,191,714	27	275,624,795	84
		32	160,087,918	16,168,879,718	33	276,570,519	86
Random sampling	b	33	290,655,390	29,356,194,390	60	278,327,625	90
		34	140,050,080	14,145,058,080	29	275,971,514	85
		Average	155,342,069	15,689,548,939	32	276,277,344	85

Table S5. Summary of the somatic mutations detected in a 36-yr-old clementine tree.

		Counts		Frequency
		Total	Assigned	
Type	**Singleton**	73	34	0.47
	Shared		39	0.53
Genetic context	**Genic**	73	20	0.27
	Intergenic		53	0.73
Genic feature	**Intron**	20	11	0.55
	CDS		9	0.45
Effect	**Non effect**	20	11	0.55
	Synonymous		4	0.20
	Non synonymous		5	0.25
Change	**Ts**	73	52	0.71
	Tv		21	0.29
Dipyrimidine site	**Yes**	73	53	0.73
	No		20	0.27
Low genic region	**Yes**	73	36	0.49
	No		37	0.51
GC > AT	**Yes**	73	35	0.48
	No		38	0.52
GC > AT - Low genic region	**Yes**	35	22	0.63
	No		13	0.37
GC > AT - Dipyrimidine site	**Yes**	35	25	0.71
	No		10	0.29

Table S6. Primers used in Sanger sequencing validations of the somatic mutations detected in a 36-yr-old clementine tree.

Variant	Forward	Reverse
6	CTCTCTGACCTTGCGCTCAT	ACAAACACCCTCTGGCACAA
25	TTAAGCCCGTACCACAACCC	ACTGATCCCCAAATGGTCGC
27	CAAGGGAGACGAACCGAACA	GAAGAGCGAGGACTGGTTCC
43	ACCAAGAAAGATGCCAACACT	TGCTAGCTTAAGTGGCAAGGA
49	TTCAGTGCATCGCCTAGAGG	ACGACCATCCAACACAACAA
51	GGCCCACTCACCAAGATTGA	TGGGGTGGAACTGACTGATT
59	CCCGAATCTACACCCACCAG	GGCCCATTTTGGTGTTCTAGC
64	TGGTGTGTTTACTTATTGAGAATGAG	AGCCACTTTAGGAAATAAGAATCTCA
72	ACATACATGTGCCTCGTAAAGGT	CGTACTTGACTTCCCGCTCA
77	GCATGCAAGGCTAGCCAAAA	AAGGATCCTTTGGTCAAGCCT
79	TTGTTTGATCCACATGACATGCA	AGCCACAGTACAGTCATGACA
81	ACAGTACCGAACCAACTAAAGCA	TGTGGGATACTCGTGTTGCA

Table S7. Sanger sequencing validations of the somatic mutations detected in a 36-yr-old clementine tree. (V, Validated; X, Not confirmed)

Sample codes					**Somatic Mutations**											
1st order branch	**2nd order branch**	**Leaf flush**		**Technical replicate**	6	25	27	43	49	51	59	64	72	77	79	81
19	6	1	a	1		V						V				
				2		V			V			V				
19	6	1	b	3				V		V						
				4												
19	0	1		5					V					V	V	
				6					V					V	V	
27	0	1		7	V		V	V	V	V	V		V		V	V
				8	V		V	V		V	V		V		V	V
36	0	1	b	9												
				10												
36	0	1	a	11										V		
				12										V		
19	6	2	a	13		V			V			V				
				14												
19	6	2	b	15				V								
				16												

27	0	2		17															
				18															
19	0	2		19					V								V		
				20					V								V		
36	0	2	a	21												V			
				22												V			
36	0	2	b	23	V		V	V		V	V		V		V	V			
				24	V		V	V		V	V		V		V	V			
19	6	3		25		V			V			V			V				
				26		V			V			V			V				
27	0	3		27	V		V	V		V	V		V		V	V			
				28	V		V	V		V	V		V		V	V			
19	0	3		29				V											
				30															
Random sampling			a	31	V		X	X		V	V		X		V	V			
				32	V		X	X		V	V		X		V	V			
Random sampling			b	33	V		V	V		V	V		V		V	V			
				34	V		V	V		V	V		V		V	V			

Table S4. Somatic mutations detected in a 36-yr-old clementine tree and their main features.

Table S8. Allele balances of the somatic mutations detected in a 36-yr-old clementine tree.

Table S9. Error estimation analysis and called sites frequency.

They are available in the online version at https://doi.org/10.1002/tpg2.20162

CHAPTER 2

CHAPTER 2

Transcriptome Analysis of the Pulp of Citrus Fruitlets Suggests that Domestication Enhanced Growth Processes and Reduced Chemical Defenses Increasing Palatability

Perez-Roman, E., Borredá, C., Tadeo, F.R., and Talon, M. (2022). Transcriptome analysis of the pulp of citrus fruitlets suggests that domestication enhanced growth processes and reduced chemical defenses increasing palatability. *Front. Plant Sci*. 13:982683. https://doi.org/10.3389/fpls.2022.982683

The student contributed to this article:

1. Collecting samples and phenotype data.
2. Conducting experimental validation.
3. Curating WGS and RNA-seq data and performing bionformatic analysis.
4. Creating and preparing visual items: plotting results, designing figures and arranging tables.
5. Reviewing and editing the original draft.

Abstract

To identify key traits brought about by citrus domestication, we have analyzed the transcriptomes of the pulp of developing fruitlets of inedible wild Ichang papeda (*C. ichangensis*), acidic Sun Chu Sha Kat mandarin (*C. reticulata*) and three palatable segregants of a cross between commercial Clementine (*C. x clementina*) and W. Murcott (*C. x reticulata*) mandarins, two pummelo/mandarin admixtures of worldwide distribution. RNA-seq comparison between the wild citrus and the ancestral sour mandarin identified 7267 differentially expressed genes, out of which 2342 were mapped to 117 KEGG pathways. From the remaining genes, a set of 2832 genes was functionally annotated and grouped into 45 user-defined categories. The data suggest that domestication promoted fundamental growth processes to the detriment of the production of chemical defenses, namely, alkaloids, terpenoids, phenylpropanoids, flavonoids, glucosinolates and cyanogenic glucosides. In the papeda, the generation of energy to support a more active secondary metabolism appears to be dependent upon upregulation of glycolysis, fatty acid degradation, Calvin cycle, oxidative phosphorylation, and ATP-citrate lyase and GABA pathways. In the acidic mandarin, downregulation of cytosolic citrate degradation was concomitant with vacuolar citrate accumulation. These changes affected nitrogen and carbon allocation in both species leading to major differences in organoleptic properties since the reduction of unpleasant secondary metabolites increases palatability while acidity reduces acceptability. The comparison between the segregants and the acidic mandarin identified 357 transcripts characterized by the occurrence in the three segregants of additional downregulation of secondary metabolites and basic structural cell wall components. The segregants also showed upregulation of genes involved in the synthesis of methyl anthranilate and furaneol, key substances of pleasant fruity aroma and flavor, and of sugar transporters relevant for sugar accumulation. Transcriptome and qPCR analysis in developing and ripe fruit of a set of genes previously associated with citric acid accumulation, demonstrated that lower acidity is linked to downregulation of these regulatory genes in the segregants. The results suggest that the transition of inedible papeda to sour mandarin implicated drastic gene expression reprograming of pivotal pathways of the primary and secondary metabolism, while palatable mandarins evolved through progressive refining of palatability properties, especially acidity.

Keywords

edible mandarins, fruit acidity, genetic admixture, secondary metabolism, wild citrus

Introduction

Our current knowledge on citrus domestication is still very unprecise (Deng et al., 2020; Rao et al., 2021; Kalita et al., 2021). During the last years, most attention has been paid to the identification of genes or gene families impacting palatability traits, fruit characteristics and reproductive behavior. Among the palatability traits, wide evidence indicates that domestication has modulated pivotal genes regulating major components of citrus flavor, such as acidity (Strazzer et al., 2019; Butelli et al., 2019; Feng et al., 2021; Borredá et al., 2022), bitterness (Chen et al., 2021) or sweetness (Li et al., 2019; Feng et al., 2021). Domestication also appears to have reduced certain chemical defenses in citrus (Gonzalez-Ibeas et al., 2021a, b; Rao et al., 2021), as reported in many crops (Köllner et al., 2008; Yactayo-Chang et al., 2020). It has also been proposed that the increased fruit size of cultivated citrus was acquired during the citrus domestication stage (Wu et al., 2018; Gonzalez-Ibeas et al., 2021b). Several reproductive characteristics closer linked to yield, have also been suggested to be key domestication targets, such as flowering (FT, TFL1 LEAFY or AP1; Rao et al., 2021), self-incompatibility (S-locus; Liang et al., 2020), and apomixis (RWP, Nakano et al., 2012; Wang et al., 2017).

In previous work, we proposed that citrus domestication was led by apomixis and hybridization phenomena (Wu et al., 2018; Wu et al., 2021), a combination that drove reticulate evolution (Hamston et al., 2018) and the formation of a syngameon (Buck and Flores-Renteria, 2022), in the genus *Citrus*. Apomixis, that gives rise to nucellar embryony (polyembryony), allows asexual reproduction of the maternal phenotype. The rise of apomixis in the ancestral mandarin lineage, provided the framework to select through clonal propagation, plants with highly appreciated organoleptic and agronomical characteristics. On the other hand, all edible citrus bear unmistakable hybridization signatures of ancestral mandarins and pummelos (Wu et al., 2014, 2018), indicating that relevant introgressed traits were selected and fixed during domestication. During the last centuries, a myriad of crosses between ancestral hybrids and admixtures, gave rise to the current basic types of edible citrus that were progressively improved through somatic mutations and recurrent selection (Talon et al., 2020). We have also

used genomic and transcriptomic analyses on wild and domesticated citrus to discriminate major determinants of evolution and domestication (Gonzalez-Ibeas et al., 2021a, b) and to identify genes involved in relevant physiological processes for domestication, (Borredá et al., 2022).

In the current work, we follow this approach comparing gene expression in the fruitlet pulp of wild inedible Ichang papeda (ICH; *C. ichangensis* Swingle), acidic Sun Chu Sha Kat mandarin (SCM; *C. reticulata*, Blanco; *C. erythrosa* Tanaka) and three palatable genetic admixtures derived from a cross between two modern commercial mandarins, namely, Clementine (CLM; *C. x clementina* Hort. ex Tanaka) and W. Murcott (WMU, *C. x reticulata* Blanco). Recent developments suggest that ICH, that is considered one of the most primitive wild forms of citrus (Swingle and Reece, 1967; Yang et al., 2017), split from the main citrus clade around 7 million years (Wu et al., 2018). The ancestor of SCM probably appeared during the last 1.4 million years, after the divergence of the two main subspecies of mainland Asian mandarins (Wu et al., 2021). The main objective of this study was to provide insights on the genetic regulation of major biological processes affected by domestication in citrus.

Material and Methods

Plant Material and Sample Processing

The plant materials used in this work were wild inedible Ichang Papeda (ICH; *C. ichangensis* Swingle), acidic Sun Chu Sha Kat mandarin (SCM; *C. reticulata*, Blanco; *C. erythrosa* Tanaka) and three palatable genetic admixtures (S1, S2 and S3) derived from a cross between two modern commercial mandarins, namely, Clementine (CLM; *C. x clementina* Hort. ex Tanaka) and W. Murcott (WMU, *C. x reticulata* Blanco). Developing fruitlets were harvested from adult trees grown under normal culture practices at the IVIA germplasm bank and experimental fields, following the protocol described in Cercós et al. (2006). In essence, homogeneous fruits were selected by uniformity of size, appearance and absence of abiotic and biotic stress symptoms. For the transcriptomic analysis, the pulp of developing fruitlets of those two pure species, ICH and SCM, and the three pummelo/mandarin admixtures, S1, S2 and S3, was collected. Fruitlets were peeled, and flavedo (exocarp) and albedo (mesocarp) discarded. The remaining tissue, the fruit flesh consisting of juice vesicles (endocarp) including the segments with their membranes and vascular bundles, was frozen under

liquid nitrogen and stored at –80ºC until analyses. Three biological replicates of each sample were taken on July, 3rd, 2020 (62 days after anthesis), at the end the cell division stage (phase I) of the development of citrus fruits (Cercós et al., 2006), a critical period for the establishment of pivotal characteristics of the citrus ripe fruits (Terol et al., 2019). A second set of pulp samples to be used in qPCR analysis determinations, consisting exclusively of juice vesicles, was collected from ripening fruits on November 21st (190 days after anthesis).

Phenotypical Data

In order to study correlation between gene expression and acidity and because acid levels do not still show accumulation at early July, fruits were also collected once a month, during the ripening period (from October to February), when the maximum acid accumulation that generally occurs in September, has already taken place (Cercós et al., 2006). Biochemical parameters (acidity, ºBrix, and maturity index) were registered in these samples. Citric acid equivalents (g/l) were determined by titration with 0.1 M sodium hydroxide and a phenolphthalein indicator. Soluble sugar content was measured with a refractometer ATAGO PR-1.

RNA Extraction, Library Preparation and Sequencing

Total RNA from pulp samples was extracted with acid phenol and precipitated with lithium chloride. Three biological replicates were used for each sample. Library preparation and sequencing were carried out by a commercial service following standard protocols. Essentially, samples enriched in mRNA were randomly fragmented and cDNA synthesized. After adapter ligation, size selection and PCR enrichment, samples were sequenced in an Illumina NovaSeq 6000 platform yielding 150 bp pair-end reads. On average, each biological replicate produced 7.04 Gb of sequence data in 23497582 raw reads.

RNA-seq Read Mapping and DEG Analysis

The *C. clementina* reference genome (Wu et al., 2014) and its annotation data, as reported at the NCBI (National Center for Biotechnology Information, 1988), were used for RNA analysis. First, raw reads were mapped using STARv2.7.6 (Dobin et al., 2013). Read counts were computed by featureCounts function in the Rsubread package (Liao et al., 2014). DESeq2 1.26 (Love et al., 2014) was used for expression analysis following

author's recommendation. Differential expressed genes (DEGs) were detected performing pair-wise comparisons of ICH and each one of segregants, against SCM. In these comparisons, the three biological replicates of each sample were treated as a group. Two combinations of Log_2 Fold Change and alpha thresholds, either $Log_2FC=0.58$ and $\alpha=0.05$ or $Log_2FC=1$ and $\alpha=0.01$, were set for expression analysis.

Pathway Analysis

The set of DEGs between ICH-SCM resulting of the comparison using the softer thresholds, $Log_2FC=0.58$ and $\alpha=0.05$, was annotated using *C. clementina* KEGG data (Kanehisa et al., 2016). Enzymatic information and KEGG identifiers of those DEGs that mapped to the set of KEGG metabolic pathways were retrieve and Pathview 4.1 (Luo and Brouwer, 2013) was used to represent the differentially expressed. For easy interpretation, differential expression was represented by three uniform colors (up, red; blue, down; yellow, undetermined) without indication of the Log_2FC. We reserved the term *undetermined* expression for genes that share the same functional annotation (i.e., same KEGG identifier), but showed opposite expression trends, as long as one of the biological replicates reached at least 100 reads. In a few cases where both the number of genes and reads with the same expression trend were unequivocally higher, the expression displayed by these genes was considered the dominant expression.

Functional Annotation and Category Assignment

In order to functionally characterize and categorize relevant DEGs found in the previous ICH-SCM comparison, a second analysis using more restrictive thresholds, $Log_2FC=1$ and $\alpha=0.01$, was performed. From the set of DEGs obtained in this second comparison, genes that were included in previous KEGG analyses and genes that did not reach 100 reads in at least one replicate, were removed. In a first step, the remaining DEGs were functionally annotated according to the Uniprot database (UniProt Consortium, 2021) and current literature. The DEGs were grouped into 45 user-defined categories, according to their involvement in major physiological, biochemical and genetic processes, and these categories were in turn split twice as much in subsequent subcategories. Genes that could not been assigned to a specific process were grouped in clusters defined by their molecular function. A group of uncharacterized DEGs, with general and ambiguous annotations or with undescribed assignments in plants was also created.

In another experiment, DEGs between the three palatable segregants and SCM were identified. Pairwise comparisons of each segregant, S1, S2 and S3, against SCM ($Log_2FC=0.58$ and $\alpha=0.05$) were parsed as above and genes that did not reach at least 100 reads in one of the biological replicates were similarly filtered out. DEGs were grouped into the aforementioned categories and common DEGs in the three segregants were included in a single list and manually characterized.

RT-qPCR Expression Analysis

RT-qPCR was used to validate RNA-seq analysis in the pulp of developing fruitlets (Supplementary Figure 1) and to test expression of target genes in juice vesicles of ripening fruits (November). Two replicates of each sample/gene combination were performed in one-step reaction in a LightCycler Instrument. Each sample was incubated with the reverse transcriptase MultiScribe (Invitrogen) at 48°C during 30 min and with an RNAse Inhibitor (Applied Biosystems). Reaction mastermix also included LightCycler FastStart DNA Master Plus SYBR Green I kit for amplification step.

Relative gene expression was calculated using ΔΔCt method. CitUBC1 (Merelo et al., 2017) and CitACTIN11 (Strazzer et el., 2019) were used as housekeeping genes for data normalization. All primer sequences used are available in Supplementary Table 1.

DNA Extraction, Sequencing and Mapping

For each segregant, a sample of fresh leaves was collected and DNA purified using CTAB extraction method. Library preparation and whole genome sequencing (WGS) were carried out by a commercial service following a standard protocol. In short, genomic DNA was randomly sheared into short fragments that, subsequently, were end-repaired, A-tailed and further ligated with Illumina adapters. Fragments with adapters were PCR amplified, size selected, and purified. Sequencing was run in an Illumina NovaSeq 6000 platform yielding 150 bp pair-ended reads. On average, each sample produced 104,484,177 raw reads, generating 31.35 Gb of sequence data. The sequences were mapped to the *C. clementina* haploid reference (Wu et al., 2014) using BWA-MEM tool (Li, 2013). Map files were sorted and indexed using Samtools (Li et al., 2009). The mean read depth of each segregant was 68.8x, 75.9x and 70.7x.

WGS data for *C. maxima* (CHP, pummelo) and *C. reticulata* (SCM, mandarin), the two pure species that make up the genome of the three segregants, were retrieved from the

Sequence Read Archive database (National Center for Biotechnology Information, 1988) and processed as above.

Variant Calling

Variant calling was performed using the GATK-4.0.0.0 software (Van der Auwera et al., 2013). We used the HaplotypeCaller tool to generate single-sample variant call format files that were combined using the CombineGVCF tool to get matrices including the samples of the study. Each site showing a quality value greater than 10 was genotyped by GenotypeGVCF and only calls tagged as SNP were filtered according to a set of standard filters specified in the variant caller practice guide. To include RNA-seq data, input files were reformatted to adapt the alignments that span introns using SplitNCigarReads tool. Raw matrices were filtered to get species informative markers (SIM). To this end, we retrieved sites holding fixed differences between *C. maxima* (CHP) and *C. reticulata* (SCM). The set of diagnostic markers only included sites with different alleles in homozygosis, supported by at least 20 reads.

Admixture Pattern

The above set of SNPs was used to define local ancestry segments along the genome of each segregant. The haplotype (pummelo or mandarin) of the admixture stretches was determined, as in Wu et al., (2018), using windows of 1000 markers. Essentially, for each SIM, the copy number of both ancestral alleles (i.e., 2, 1 or 0) was recorded and the ancestry of the window was inferred from the most frequent one. Stable segments were considered when a minimum of five windows in a row exhibited the same inheritance. Otherwise, the ancestry of the nearest stable block was applied. These data were used to compute the distribution of expressed genes in each segregant. Only genes that were covered with a minimum of 10 reads in any of the biological replicates were considered. Haplotype combinations were named, MA/MA (mandarin/mandarin), PU/MA (pummelo/mandarin), and PU/PU (pummelo/pummelo).

Allele Differential Expression

SIMs were additionally used to study differentially expressed alleles in PU/MA regions. Only genes with heterozygous sites covered by a minimum average depth of 10x, showing identical genotype in the three biological replicates, were initially considered for this analysis. The final set of cDNAs used in the analysis of allele differential

expression included genes, in which approximately the 80 % of the SNPs spanning their sequences met these conditions. From this set of target genes, the occurrence of homozygous SIMs was assessed as exclusive expression of a certain allele.

Admixture Validation

We used PCR and Sanger sequencing to validate both a sequence change causing an admixture pattern shift such as crossover, and an event of allele differential expression (Supplementary Figures 2 and 3).

Results

To study differential gene expression as related to domestication in the pulp of developing fruitlets, two independent comparisons were performed. In one of them, the transcriptome of inedible Ichang papeda (ICH) and acidic Sun Chu Sha Kat mandarin (SCM) were contrasted. In a second evaluation, this mandarin was compared with a group formed by three palatable genetic pummelo/mandarin admixtures derived from a cross between the commercial mandarins Clementine and W. Murcott.

Gene Expression in Inedible Wild Ichang Papeda versus Acidic Sun Chu Sha Kat Mandarin

The transcriptome comparison between the developing fruitlet pulp of ICH and SCM (Log_2 Fold Change threshold of 0.58 and alpha of 0.05) identified 7267 DEGs (Supplementary Table 2). Out of this set of genes, 2342 were mapped to 117 pathways, mostly related to metabolism, genetic information and cellular processes (Supplementary Figures 4.001-4.117), as defined in the KEGG database resource (Kanehisa et al., 2016). Genes that did not map to the KEGG collection, were compared using more stringent conditions (Log_2 Fold Change threshold of 1 and alpha of 0.01). This analysis rendered 2832 DEGs (Supplementary Table 3) that were, first, functionally annotated according to the Uniprot database (UniProt Consortium, 2021), and then grouped into 45 user-defined categories. Figure 1 presents the 32 categories encompassing more than 10 members, excluding the group of uncharacterized processes. This classification overall indicates that only categories included in Secondary Metabolism and categories related to protein metabolism (Translation, Ubiquitination, Trafficking) were upregulated in the papeda while most of those grouped in Development, Regulation and Transport, Signaling, Gene Expression,

Cellular Growth and Stress responses were downregulated. Thus, more than 60% of those 2832 DEGs were downregulated in the fruit flesh of ICH, a percentage rather identical to that observed in the group of uncharacterized genes. The results described below are restricted to categories potentially involved in domestication while the description and comments affecting to the rest of results are available in Supplementary Material.

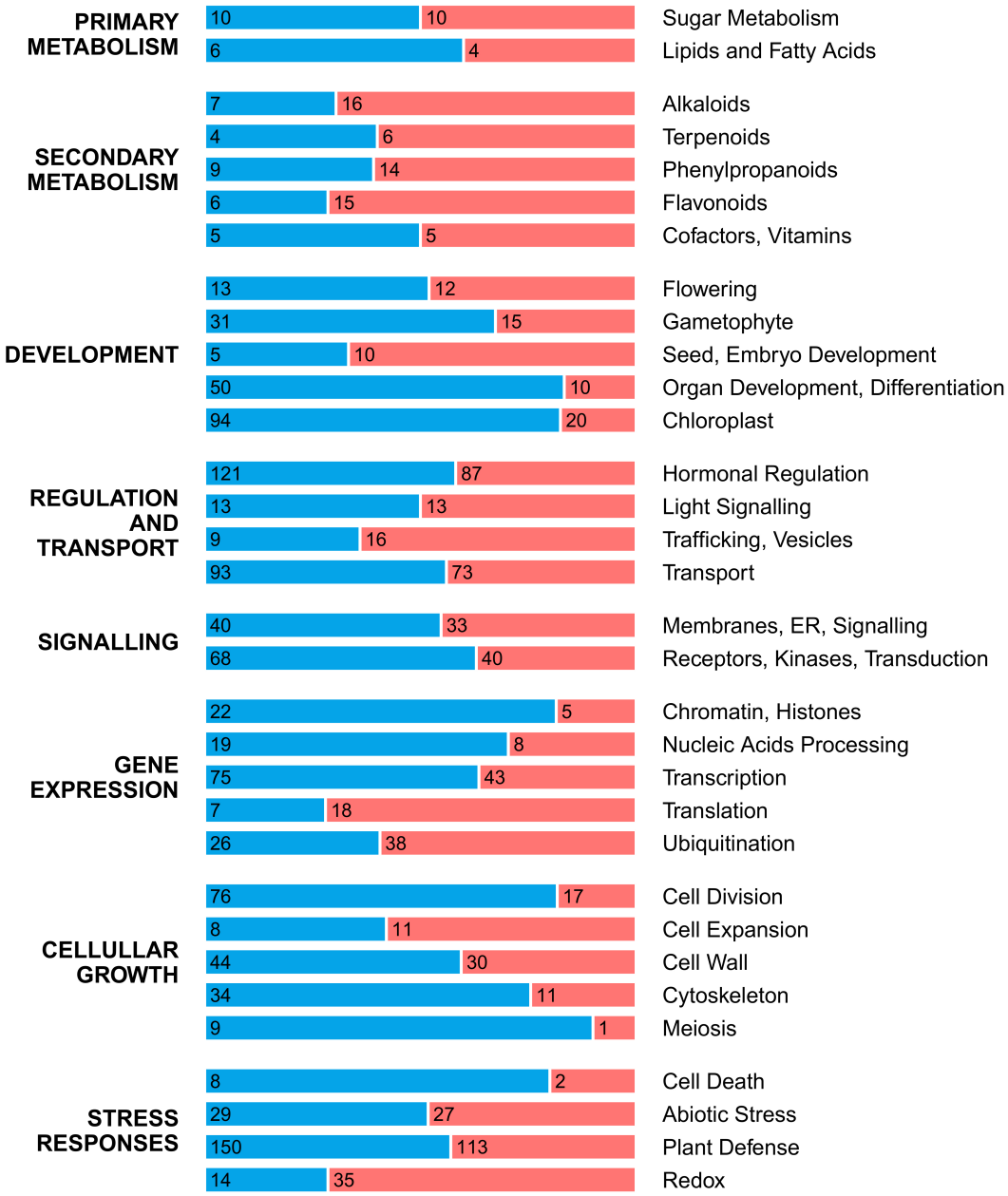


Figure 1. Distribution of up- and downregulated DEGs in the pulp of developing fruitlets of ICH as related to SCM. DEGs were manually classified into 8 major groups that were subsequently split in 32 gene categories corresponding to the most populated categories presented in Supplementary Table 3, excluding the group of uncharacterized

genes. Length of red and blue stretches represents the frequency of up- and downregulated genes, respectively, while numbers refer to the number of genes included in each category.

Alkaloids. We grouped 23 DEGs into the Alkaloids category (Supplementary Table 3) while the KEGG pathway mapping added several additional genes to this cluster. According to the expression levels of transcripts associated with the synthesis of several principal alkaloids, the data suggest that these compounds are generally upregulated in ICH. In this species, for instance, caffeine synthase is upregulated (LOC112100740). The transcripts of codeine 3-O-demethylase, the last step in the biosynthesis of morphine (LOC18036840) in the Isoquinoline Alkaloid Pathway, also was upregulated. Similarly, synthesis of major indol alkaloids, i.e., ajmaline, vinblastine and vincristine and also that of iridoid compounds appears to be favored as suggested by the upregulation of relevant biosynthetic steps controlling the conversions of monoterpenoid precursors of indole alkaloid biosynthesis. These upregulated steps are geraniol 8-hydroxylase (5 genes out of 6), 8-hydroxygeraniol dehydrogenase (LOC18037214), the dehydrogenase involved in the biosynthesis of oxogeranial from hydroxygeraniol and 7-deoxyloganetin glucosyltransferase (LOC18042369), an iridoid glucosyltransferase involved in the synthesis of secologanin, one of the major intermediates in the indole alkaloid biosynthesis. It was also observed that a vinorine synthase (LOC18037906), a gene coding for the acetyltransferase catalyzing the formation of vinorine, a precursor of the monoterpenoid indole alkaloid ajmaline, was also upregulated. The synthesis of nicotine, a major alkaloid derived from nicotinic acid (Supplementary Figure 4.067), however, does not appear to be activated. In the acridone alkaloid biosynthesis, two methanol O-anthraniloyltransferases genes (LOC18050765 and LOC18050764), in principle implicated in the synthesis of methyl anthranilate, showed opposite expression tendencies. Likewise, evidence for the upregulation of the tropane biosynthesis, leading to alkaloids such as atropine, hyoscyamine and scopolamine was not obtained. The analysis of gene expression certainly showed upregulation of two tropinone reductase-like genes (LOC18039952 and LOC18039951) but the corresponding proteins do not appear to exhibit tropinone reductase activity (Jirschitzka et al., 2012). Interestingly, synthesis of both dopamine and serotonin (Supplementary Figures 4.089 and 4.033), two amides that are considered bioactive alkaloid neurotransmitters, in principle, appear to be downregulated in ICH, since tyrosine/DOPA decarboxylase 5 (aromatic L-amino acid decarboxylase; LOC18043348,

Supplementary Table 3, and LOC18046227, EC:4.1.1.28, Supplementary Figures 4.033 and 4.087, Supplementary Table 2) is repressed in this species. This enzyme (EC:4.1.1.28) may render tyramine, tryptamine, dopamine and serotonin. Likewise, a CYP71P1 gene (LOC18041681, EC:1.14.-.-, Supplementary Figure 4.033, Supplementary Table 2) encoding tryptamine 5-hydroxylase that also catalyzes the conversion of tryptamine to serotonin, is repressed too.

Phenylpropanoids. The last steps in the formation of the derived aldehyde and alcohol of major phenylpropanoids (Supplementary Figure 4.081, Supplementary Table 2), such as cinnamic, coumaric, caffeic, ferulic, hydroxyferulic and sinapic acids, are predominantly upregulated in ICH, as observed for instance, cinnamoyl-CoA reductase (LOC18035881, EC:1.2.1.44). However, trans-cinnamate 4-monooxygenase, CYP73A, (LOC18055509, EC:1.14.14.91) and 4-coumarate--CoA ligase (LOC18034975, LOC18050151 and LOC18036308, EC:6.2.1.12) the main steps in the synthesis of p-Coumaroyl-Co A, the precursor of flavonoids and the non-flavonoid polyphenols, stilbenoid, diarylheptanoic and gingerol biosynthesis, are downregulated. In this pathway, trans-resveratrol di-O-methyltransferase (LOC18051743 and LOC18031586, Supplementary Tables 2 and 3), the last step in the biosynthesis of the antifungal phytoalexin pterostilbene, is expressed a higher level in the papeda. In Supplementary Table 3, there are listed other DEGs that participate in the regulation of the phenylpropanoid biosynthesis such several members of the Cytochrome P450 71A1 family, that appear to act as trans-cinnamic acid 4-hydrolases, or the MYB family transcription factor PHL11. The enzymes listed in this table, caffeic acid 3-O-methyltransferases, caffeoylshikimate esterases-like, cinnamoyl-CoA reductases, and shikimate O-hydroxycinnamoyltransferases, that are not mapped at the KEGG pathways, appear in principle to be associated with the phenylpropanoid pathway and at least some of them, with lignin biosynthesis.

Flavonoids. In the flavonoid pathway, the synthesis of the pivotal intermediate flavanone naringenin in ICH appears to be downregulated (Supplementary Figure 4.082, Supplementary Table 2), since chalcone isomerase (LOC18044429, EC:5.5.1.6), was repressed, although there were at least four different chalcone synthetases (LOC18042808, LOC18042812, LOC18033130 and LOC18051925, E.C:2.3.1.74), with opposite genetic expression levels. The synthesis of isoflavonoids (Supplementary Table 3) was mostly characterized by the upregulation of 2-hydroxyisoflavanone

dehydratase-like (LOC112096719 and LOC112098436), that catalyze the final step in the formation of the isoflavonoid skeleton rendering daidzein, of isoflavone 4'-O-methyltransferase (LOC18053393) involved in the biosynthesis of formononetinin and of isoflavone 2'-hydroxylases (LOC18043085), that mediates the hydroxylation of daidzein and formononetin, to yield 2'-hydroxyisoflavones. In spite of downregulation of naringenin, the formation of the polyphenolic flavonols, kaempferol, quercetin and myricetin appears to be upregulated (Supplementary Figure 5). Supplementary Table 3, for instance, reports several flavonoid 3'-monooxygenases (LOC18053376,) and flavonoid 3',5'-hydroxylases (LOC18048580) controlling the conversion of naringenin to eriodictyol, dihydrokaempferol, dihydroquercetin and dihydromyricetin, and Supplementary Figure 4.082 shows upregulation of naringenin 3-dioxygenase (LOC18036490, EC:1.14.11.9) and flavonol synthase (LOC18037475, EC:1.14.20.6, Supplementary Table 2), two regulating enzymes of the synthesis of kaempferol, quercetin and myricetin. The synthesis of the anthocyanidins, pelargonidin, cyanidin and delphinidin, and their corresponding anthocyanins (anthocyanidin glycosides; Khoo et al., 2017) was similarly upregulated in the papeda. Supplementary Table 2, for example, shows that two limiting steps in the synthesis of anthocyanidins and anthocyanins, anthocyanidin synthase (LOC18047155, EC:1.14.20.4, Supplementary Figure 4.082) and anthocyanidin 3-O-glucosyltransferase (LOC18047244, EC:2.4.1.115, Supplementary Figure 4.083), respectively, were clearly upregulated. In addition, Supplementary Table 3 enumerates a number of upregulated members of several gene families, coumaroyl-CoA:anthocyanidin 3-O-glucoside-6"-O-coumaroyltransferase 1 (LOC18050842 and LOC18032737) malonyl-CoA:anthocyanidin 5-O-glucoside-6"-O-malonyltransferase (LOC18055666, LOC18038126, LOC18044783 and LOC18046030) and putative anthocyanidin reductase (LOC18047966), suggesting that the metabolism of anthocyanins is very active in this species. It should be mentioned, that mandarins do not contain anthocyanins likely because the Ruby gene (synonymous AN2) is not functional in these varieties (Butelli et al., 2019; Wu et al., 2021). Supplementary Table 3 also lists 2 flavonol-specific transcription activators, MYB11 and MYB111, (LOC18049115 and LOC18031574) involved in the regulation of several genes of the flavonoid biosynthesis. The synthesis of the phytoalexin glyceollin, however, does not appear to be promoted since 3,9-dihydroxypterocarpan 6A-monooxygenase (LOC18031687), a previous biosynthetic step, was downregulated.

Terpenoids. In the terpenoid pathway, there were clear-cut differences between both species. The papeda showed upregulation of pivotal genes of the mevalonate pathway, in detriment of the MEP pathway, that takes place in plastids (Supplementary Figure 4.072). Thus, as related to the biosynthetic regulation of the isoprenoid precursors, isopentenyl pyrophosphate was promoted against dimethylallyl pyrophosphate. In addition, the pivotal gene, geranylgeranyl diphosphate synthase, (dimethylallyltranstransferase, LOC18039078, EC:2.5.1.1, and LOC18039079, EC:2.5.1.29, Supplementary Table 2), giving rise to main precursors of the different terpenoid types, were also expressed at higher levels. The synthesis of monoterpenoids was not basically modified except for the upregulation of 2 out of 3 (R)-limonene synthases 1 (LOC112098486 and LOC112098571, Supplementary Table 3), catalyzing the conversion of geranyl diphosphate to (+)-(4R)-limonene. In the diterpenoid biosynthesis, trimethyltridecatetraene/dimethylnonatriene synthase, CYP82G1, (LOC18049179 and LOC18049178, EC:1.14.14.58, Supplementary Figure 4.074, Supplementary Table 2), catalyzing the production of the volatile homoterpenes DMNT and TMTT, was also upregulated. In the sesquiterpenoid pathway, the synthesis of farnesene appears to be downregulated since transcripts of α-farnesene synthase (LOC18053589, EC:4.2.3.46, Supplementary Figure 4.078, Supplementary Table 2) are present at lower levels, while there were two (3S,6E)-nerolidol synthase 1 genes (LOC18033168 and LOC18051782, EC:4.2.3.48) expressed in opposite directions. Expression of premnaspirodiene oxygenase (LOC18052043), involved in the biosynthesis of the sesquiterpenoid, solavetivone, a potent antifungal phytoalexin, was upregulated, while that of α-copaene synthase-like (LOC112095677) that converts farnesyl diphosphate to the bicyclic olefins α-copaene and (E)-β-caryophyllene and participates in the synthesis of the macrocyclic sesquiterpene germacrene D, was downregulated (Supplementary Table 3). One important gene implicated in the synthesis of steroids and triterpenoids was squalene monooxygenase, SQLE, (LOC18033947 and LOC18033838, EC:1.14.14.17, Supplementary Figures 4.012 4.078, Supplementary Table 2), that was downregulated in the papeda and therefore, probably limiting the flux toward these compounds. Consistently, the number of DEGs regulating triterpenoid metabolism were also scarce, since only the synthesis of β-amyrin appears to be upregulated (LOC18034001, EC:5.4.99.39, LOC18045727 and LOC18053646, Supplementary Tables 2 and 3). Regarding carotenoids, only transcripts coding for enzymes mediating phytoene synthesis (LOC18051922, EC:2.5.1.32,

Supplementary Figure 4.076, Supplementary Table 2) were upregulated in the papeda, while the metabolism of α- and β-carotenes, including the synthesis of cryptoxanthin, lutein, astaxanthin and the xanthophyll cycle, was strongly repressed. Supplementary Table 3 also lists two upregulated β-D-glucosyl crocetin β-1,6-glucosyltransferases (LOC18045744 and LOC18044391), catalyzing the β 1-6 glucosylation of crocetin, a natural apocarotenoid. In addition, this table reports on other two glycosyltransferases (LOC18054015 and LOC18033356) conjugating diterpenes that are downregulated, and an upregulated gene of the diterpenoid metabolism, cytochrome P450 76M5 (LOC18031421), involved in the biosynthesis of oryzalexin, a class of phytoalexins. The terpenoid pathways implicating plant hormones such as GAs, ABA, cytokinins and brassinosteroids are described in the Hormonal category. Upregulation of the committed steps of the synthesis of tocopherol and tocotrienol (vitamin E), homogentisate phytyltransferase/ geranylgeranyltransferase (LOC18055996, EC:2.5.1.115 and EC:2.5.1.116, Supplementary Figure 4.013, Supplementary Table 2) was also found in the papeda.

Hormonal Regulation. The comparison between the transcriptome of ICH and SCM also rendered a relatively high number of DEGs involved in hormone biosynthesis and action. According to the KEGG mapping of biosynthetic DEGs (Supplementary Table 2), the synthesis of active cytokinins, brassinosteroids and ethylene was downregulated in the papeda. Regarding cytokinins biosynthesis, transcripts for cytokinin dehydrogenase, CKX, (LOC18042746 and LOC18033392, EC:1.5.99.12), an enzyme that inactivates isopentenyl adenine, were upregulated, in contrast to those of two glucosyltransferases conjugating zeatin, (LOC18031300, EC:2.4.1.215, and LOC18038024 and LOC18037288, EC:2.4.1.-), that were repressed (Supplementary Figure 4.077). The synthesis of brassinosteroids depending upon steroid precursors, also appears to be strongly downregulated (Supplementary Figure 6), since most steps, including last steps in the synthesis of brassinolide, such as brassinosteroid-6-oxidase 1, CYP85A1, (LOC18038016 and LOC18044268, EC:1.14.-.-) and PHYB activation tagged suppressor 1, CYP734A1, (LOC18054829, EC:1.14.-.-), were downregulated (Supplementary Figure 4.075). The generation of ethylene does not appear to be promoted either in the papeda, because the last step in its synthesis, 1-aminocyclopropane-1-carboxylate oxidase (LOC18050524, EC:1.14.17.4) was repressed, although the previous conversion catalyzed by 1-aminocyclopropane-1-

carboxylate synthase (LOC18048242 and LOC18046436, EC:4.4.1.14), was upregulated (Supplementary Figure 4.024). Although there was upregulation of early steps in the gibberellin biosynthesis, including the conversions between ent-kaurene to inactive GA_{12} (LOC18046916, EC:1.14.14.107, Supplementary Figure 4.074) no differences in the expression of biosynthetic genes controlling the formation of active GAs between both species were found. A similar situation was observed in the synthesis of jasmonate and methyl-jasmonate, characterized by the upregulation of no less than 6 early biosynthetic steps (MFP2, LOC18032845, EC:4.2.1.17 and ACX, LOC18047109, EC:1.3.3.6, Supplementary Figure 4.053). The synthesis of xanthoxin, a precursor of ABA, and ABA degradation, was upregulated in the papeda, since genes coding for 9-cis-epoxycarotenoid dioxygenase (LOC18043465 and LOC18050641, EC:1.13.11.51) and abscisic acid 8'-hydroxylase (LOC18039758, EC:1.14.14.137, Supplementary Figure 4.076), the proteins controlling these conversions were expressed at higher levels. The data also suggest that auxin synthesis was also promoted because two amidases (LOC18033993 and LOC18034584, EC:3.5.1.4), an aldehyde dehydrogenase (LOC18036436, EC:1.2.1.3) and at least an indole-3-pyruvate monooxygenase (LOC18032700, EC:1.14.13.168) participating in the auxin synthesis, were expressed at higher levels (Supplementary Figure 4.033). In contrast, the amino acid derived polyamines, spermidine (LOC18039571, EC:1.5.3.17), spermine (LOC18043803, LOC18051913 and LOC18054425, EC:1.5.3.16) and putrescine (LOC18049691, EC:4.1.1.17) were apparently downregulated (Supplementary Figure 4.029; Supplementary Table 4). Moreover, the expression levels of pivotal receptors, transporters and regulators implicated in the hormone signal transduction, indicate that the receptors of cytokinins, CRE1 (LOC18052080, EC:2.7.13.3), brassinosteroids, BRI1 (LOC18035850, EC:2.7.10.1 and EC:2.7.11.1) and ethylene, ETR (LOC18031847, EC:2.7.13.-) are repressed in the papeda, like most components of the auxin transduction pathway, including the auxin flux carrier AUX1 (LOC18034947, LOC18038157 and LOC18045480) and the regulators TR1 (LOC18052162) or GH3 (LOC18035901, LOC18053209, LOC18054772, LOC18033692 and LOC18041056). In contrast, JAR1 (LOC18049830, EC:6.3.2.52, ST 7267) and PP2C (LOC18043434, EC:3.1.3.16), major regulators of jasmonate and ABA responses, respectively, and GID1 (LOC18049839 and LOC18043172), the receptor of GAs, were both upregulated (Supplementary Figure 4.109). Regarding MAPK Signaling Pathway, upregulation of MKK3 (LOC18051058, EC:2.7.12.2) and MPK6 (LOC18047683, EC:2.7.11.24,

Supplementary Figure 4.107, Supplementary Table 2), was the most significant observation as related to hormonal regulation.

In addition, Hormonal Regulation category included 208 DEGs, out of which more than 58% were downregulated in ICH (Supplementary Table 3). Although all phytohormones were represented in this set of genes, not all of them exhibited the same down/up regulation ratio. In particular, transcripts related to auxins (49/19, transport, homeostasis, ARFs, AUX/IAA, SAURs, response, biosynthesis, signaling), cytokinins (6/2, transport, receptors, transcription factors), gibberellins (13/4, biosynthesis, response, signaling, transcription factors) and jasmonic acid (13/8, transcripts related to biosynthesis, transport, response, receptor, induced response, signaling, transcription factors) had higher number of downregulated genes. Downregulation frequency of these categories ranked from 0.62 to 0.76. Genes linked to ethylene (6/18, transcription factors, induced responses, ERFs) displayed the opposite tendency, while transcripts associated with ABA (21/20, biosynthesis, response, receptor, induced response, signaling, transcription factors), polyamines (1/1, transporters), salicylic acid (8/10, transcription factors, induced responses, signaling, biosynthesis) and brassinosteroids (4/4, transcription factors, response, signaling, homeostasis) exhibited a down/up ratio that hardly departs from 0.5 (Supplementary Table 3).

Other categories related to growth, such as **Gametophyte, Organ Development, Differentiation, Chloroplast, Cell division, Meiosis, Cytoskeleton, Cell Wall, Receptors and Protein Kinases, Chromatin, Histones, Transcription, and Nucleic Acids Processing** were also downregulated in the papeda (Supplementary Material).

Primary Metabolism. Regarding carbon metabolism, striking differences were found between both species, since practically all genes coding enzymes of central regulatory pathways such as glycolysis (Supplementary Figure 4.001), including the generation of pyruvate, acetyl CoA and acetaldehyde (Supplementary Figure 4.057), were clearly up-regulated in ICH. In the pentose phosphate pathway (Supplementary Figure 4.003), the synthesis of the pivotal intermediate, glyceraldehyde 3P, was similarly upregulated. Expression of genes involved in sucrose synthesis and degradation do not appear to be clearly modified, while the formation of ADP-glucose and amylose (Supplementary Figure 4.042), but not that of starch, also were upregulated in the papeda, as that of α amylase (LOC18043125 and LOC18045113, EC:3.2.1.1, Supplementary Figure 4.042, Supplementary Table 2). In this species, it was also repressed a regulatory subunit of the

probable trimeric SNF1-related protein kinase, (SnRK; LOC18052199, Supplementary Table 3) complex, that appears to play a role in the transduction cascade regulating gene expression and carbohydrate metabolism. In the papeda, other important differences were found in the metabolism of organic acids, especially the tricarboxylic cycle (TCA), that showed up-regulation of most genes coding for their regulatory enzymes (Supplementary Figure 4.002, Supplementary Table 2), including those of the pyruvate dehydrogenase complex (LOC18045003 and LOC18037469, EC:2.3.1.12). A noticeable exception to this observation, however, was the conversion of oxoglutarate to succinyl-CoA that was down-regulated (LOC18044474, EC:1.2.4.2). Succinyl-CoA synthetase alpha subunit (LOC18045656, EC:6.2.1.4 and EC:6.2.1.5), on the other hand, was upregulated. Expression of genes regulating cytoplasmatic organic acid metabolism was similarly altered, since several subunits of ATP-citrate lyase, ACLY (LOC18032750 and LOC18043354, Supplementary Table 2 and LOC18039980, Supplementary Table 3), that converts citrate into oxaloacetate and cytosolic acetyl-CoA, were upregulated. Likewise, genes coding for a series of enzymes acting sequentially, such as aconitate hydratase, ACO3 (LOC18055416, EC:4.2.1.3, Supplementary Figure 4.058), one isocitrate dehydrogenase [NADP], IDH1, (LOC18031748, EC:1.1.4.2) and aspartate transaminase, GOT1, (LOC18054901, EC:2.6.1.1, Supplementary Figure 4.017), rendering the amino acid glutamate from 2-oxogluarate, were also upregulated. In addition, three additional enzymes, glutamate decarboxylase, GAD5, (LOC18046053 and LOC18052541, EC:4.1.1.15), butyrate pyruvate transaminase, POP2, (LOC18039191, EC:2.6.1.96) and succinate-semialdehyde dehydrogenase, SSADH, (LOC18031917, EC:1.2.1.24), that together make up the GABA shunt, showed the same tendency (Supplementary Figure 4.021). The conversion of glutamate to glutamine (LOC18044424, EC:6.3.1.2) and pyrroline-carboxylate (LOC18045924, EC:1.2.1.88) also was favored in ICH (Supplementary Figure 4.021). The oxidative phosphorylation likewise seems to be more active in ICH, since all DEGs implicated in this process were upregulated (Supplementary Figure 4.014, Supplementary Table 2). These genes code for several components of complex I, NADH hydrogenase, including NADH-quinone oxidoreductase subunit A (LOC18034396, EC:7.1.1.2); complex III, cytochrome c reductase, including ubiquinol-cytochrome c reductase cytochrome b/c1 subunit (LOC18051936, EC:7.1.1.8), complex IV, cytochrome oxidase, including cytochrome c oxidase cbb3-type subunit I, COX6A and COX6B, (LOC18037730 and LOC18055702, EC:7.1.1.9)

and complex V, ATP synthase, including H+-transporting ATPase (LOC18055993, LOC18039766, LOC18053876 and LOC18035736, EC:7.1.2.1). Major changes in the lipids and fatty acids pathways are discussed in Supplementary Material.

Amino acid metabolism. Amino acid metabolism and the synthesis of several derived compounds differ in both species. Supplementary Table 4 speculates on the regulation of the synthesis of these compounds in each species, based on gene expression levels of the last regulatory steps (Supplementary Figures 4.017, 4.021-4.036, 4.039 and 4.042). According to this information, the synthesis of amino acids tends to be upregulated in ICH, except for the production of valine, leucine and isoleucine that was clearly repressed, and their degradation upregulated. Data related to amino acid-derived hormones and alkaloids are specified elsewhere in this section. In the cyanoamino acid metabolism, it is worth to mention that in ICH, linamarin synthase (LOC18036876, Supplementary Table 3), an UDP glycosyltransferase producing cyanogenic glucosides was upregulated. Consistently, β-glucosidase 13 (orthologous of β-glucosidase 12 of cassava, linamerase, LOC18044658, EC:3.2.1.21, Supplementary Table 2), that converts cyanogenic glucosides into acetone cyanohydrins such as mandelonitrile, and mandelonitrile lyase (LOC18037363, EC:4.1.2.10), that releases HCN, hydrogen cyanide, from the acetone were both downregulated (Supplementary Figure 4.039). The synthesis of glucosinolates may be promoted in the papeda, since a flavin-containing monooxygenase FMO GS-OX-like 4 (LOC18049889, Supplementary Table 3) and two mRNAs coding for cytochrome P450 83B1 (LOC18041363 and LOC18046366, ST 2832) that catalyze the oxime metabolizing step in indole glucosinolate biosynthesis were upregulated.

Transport. In the papeda, the Transport category was also enriched in downregulated genes (Figure 1), although specific differences were found among the wide range of transport systems included in this group (Supplementary Table 3). For instance, ABC transporters for glutathione S-conjugates, vacuolar ATPases, and copper, magnesium, sulfate, and zinc transporters were mostly upregulated, while transporters of amino acids, ascorbate, cation channels, proton antiporters, components of the mitochondrial electron transport chain, mechanosensitive channels, metal-nicotianamine transporters, nitrate, sodium, cadmium, nuclear import, oligopeptides and xenobiotics were mostly downregulated. Other transporters, such as aquaporins, purines, and transporters of boron, calcium, and potassium showed similar number of up- and downregulated genes.

The number of sugar transporters was relatively high (19) and some genes were highly expressed in SCM mandarin in contrast to ICH, such as the vacuolar hexose transporter SWEET17 (LOC18032835) and other hexoses carriers (LOC18031330 and LOC18048094) or monosaccharide transporters (LOC18031593). ALTM4 (LOC18043583), an aluminum-activated malate transporter was also downregulated in papeda.

Abiotic stress. As related to stress responses, upregulation enrichment was observed in the papeda fruitlet flesh predominantly in three cases (Supplementary Table 3): in heat shock proteins (4/11), in cold responses (3/5), and in oxidative stress (6/35). Out of the 5 upregulated genes involved in cold responses, CORA, a cold and drought-regulated protein, showed the highest expression observed in this analysis (LOC18054952). Among the 3 downregulated transcripts, there were two transcription factors, the activator ICE1 (LOC18051975) and the repressor MYBS3 (LOC18055469), that regulate the cold-induced transcription of DREB1/CBF genes. Additional genes involved in cold responses mediated by ABA were also differentially expressed among both species. For instance, negative regulators of ABA such as MSI4 (LOC18036552) and ERD15 (LOC18045483) were expressed at lower levels. JUB1 (LOC18044819), another gene participating in the response to freezing was upregulated. As mentioned before, a further category enriched in the papeda with upregulated DEGs was that of Redox (Figure 1, Supplementary Table 3). It includes mostly genes coding for oxidoreductase enzymes, that play major roles in the antioxidant defense system, such as oxidases, reductases, peroxidases, cytochromes P450, mono- and dioxygenases and glutathione S-transferases. In the glutathione-ascorbate cycle, an efficient metabolic pathway to detoxify H_2O_2, monodehydroascorbate reductase (LOC18039033, E.C:1.6.5.4) was upregulated, while ascorbate peroxidase (LOC18039197, LOC18040244, LOC18037637, LOC18042802 and LOC18035392, E.C. 1.11.1.11,) appears to be predominantly repressed (Supplementary Figures 4.007 and 4.041, Supplementary Table 2). Two glutathione peroxidases (LOC18047364, EC:1.11.1.9, Supplementary Figure 4.041, Supplementary Table 2 and LOC18047405, Supplementary Table 3), that were also upregulated, might further contribute to H_2O_2 removal. In addition, the data indicate that several sequential enzymes, such as nicotinamidase (LOC18034819, Supplementary Table 3), nicotinate phosphoribosyltransferase 1 (LOC18036859, EC:6.3.4.21), pyrimidine and pyridine-

specific 5'-nucleotidase (LOC18050280, EC:3.1.3.-), nicotinamide/nicotinic acid mononucleotide adenylyltransferase (LOC18055145, EC:2.7.7.1), and NAD^+ kinase (LOC18042070, EC:2.7.1.23,), implicated through the salvage pathway in the synthesis of the central coenzyme NAD^+/NADH, were upregulated in the papeda (Supplementary Figure 4.067, Supplementary Table 2).

Gene Expression in Palatable Pummelo/Mandarin Genetic Admixtures versus Acidic Sun Chu Sha Kat Mandarin

To select mRNAs that were differentially expressed in the pulp of developing fruitlets of acidic and palatable mandarins, the transcriptomes of SCM and three palatable pummelo/mandarin genetic admixtures, coded S1, S2, and S3, were compared through RNA-seq analysis. Pairwise comparisons of each segregant against SCM (Log_2FC = 0.58 and α = 0.05) were generated and the 357 DEGs showing similar expression patterns in the three palatable mandarins and opposite in SCM were selected following the criteria utilized in previous comparisons (Supplementary Table 5).

Alkaloids. Out of the 8 DEGs clustered in the Alkaloids category, 7 of them belonged to three groups that were previously identified in the comparison between ICH and SCM. Five of these genes (3/2) belonged to the group of tropinone reductase homologs, that as commented above do not appear to possess tropinone reductase activity. LOC18039754 and LOC112096160, the 2 genes with highest expression in this group showed opposite tendencies. The other 2 genes, geraniol 8-hydroxylase (LOC18040911), involved in the biosynthesis of terpenoid indole alkaloids and in the biosynthesis of flavonoids, and methanol O-anthraniloyltransferase (LOC18050771), that generates methyl anthranilate, were both upregulated in the segregants. The remaining gene was downregulated and is annotated as hyoscyamine 6-dioxygenase-like (LOC112100197), the limiting step in the synthesis of scopolamine in the tropane alkaloid biosynthesis.

Terpenoids. The Terpenoids category (4/2) was characterized by the upregulation of zeta-carotene desaturase (LOC112098137), chloroplastic/chromoplastic-like, that plays a crucial role catalyzing the conversion of zeta-carotene to lycopene in the biosynthesis of carotenoids. The three palatable mandarins also showed upregulation of (-)-germacrene D synthase-like (LOC112100727), suggesting an activation of the synthesis of the germacrin-type sesquiterpenoid, germacrene D, a class of volatile organic

hydrocarbon with antimicrobial and insecticidal properties. The conversion of these sesquiterpenes to the derived lactones was probably downregulated, since α-copaene synthase-like (LOC112095677), that catalyzes several of these conversions, was repressed. In the acyclic sesquiterpenoid pathway, two additional downregulated genes were (E)-β-farnesene synthase-like (LOC112101583), a cyclase catalyzing the production of β-farnesene, and dimethylnonatriene synthase (LOC18049179), a cytochrome P450 82G1 involved in the biosynthesis of homoterpenes such as TMTT. (R)-limonene synthase 1 (LOC112098571), chloroplastic-like, that synthetizes the monoterpene limonene also was downregulated.

Sugar Metabolism. In the three palatable mandarins, sugar metabolism (6/1) was characterized by the repression of sucrose synthase 2 (LOC18032959), probable galactinol-sucrose galactosyltransferase 2 (LOC18031328) and stachyose synthase (LOC18050124). The first gene is implicated in sucrose cleavage and the other two in the synthesis of raffinose, stachyose and verbascose. Other repressed genes were probable trehalose-phosphate phosphatase F (LOC18049897), that produces free trehalose and phosphoenolpyruvate carboxylase kinase 1, that through decarboxylation renders oxalacetate to fuel the citric acid cycle. The only gene upregulated in this category was enolase (LOC18031514), that is responsible of the conversion of 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP).

Transport. This category grouped a set of genes (10/9) showing different patterns of expression. Relevant upregulated genes, for instance, were an aquaporin (LOC18045515), a calcium-transporting ATPase (LOC18042874), a potassium transporter (LOC18043735), and two different sugar transporters ERD6-like (LOC18040371 and LOC18046355), which are vacuolar H^+/glucose symporters involved in the export of glucose to cytosol (Klemens et al., 2014). Downregulated genes were a boron transporter (LOC18036378), a nitrate reductase (LOC18041814), a zinc transporter (LOC18050857), and overall, an ATPase 10, plasma membrane-type (LOC18035736), that has been previously associated with citric acid accumulation in lemon juice (Aprile et al., 2011). There were also two members of the sodium/hydrogen exchanger gene family (LOC18040436 and LOC18039966), involved in acidity regulation in oranges (Wang et al., 2021), expressed in opposite directions.

Other categories are discussed in Supplementary Material, although two transcripts should be mentioned for their potential relevance. One of these is UDP-

glycosyltransferase 74B1 (LOC18044914), that is involved in the biosynthesis of benzyl-glucosinolate and appears to be downregulated in the three segregants. The other gene is 2-methylene-furan-3-one reductase (LOC18050984), that codes for the enone oxidoreductase rendering furaneol, the key flavor compound in strawberries (Raab et al., 2006), and is expressed at higher levels in the palatable mandarins.

Gene Expression as Related to Acidity in Palatable Pummelo/Mandarin Genetic Admixtures versus Sun Chu Sha Kat Mandarin

The finding that ATPase 10, plasma membrane-type, (Aprile et al., 2011), a pivotal component of the vacuolar proton-pumping P-ATPase complex that regulates acidity in citrus (Strazzer et al. 2019), was downregulated in the three segregants, prompted us to focus our attention on the gene expression of the rest of components of this complex and of other related genes previously associated with this process (Huang et al., 2021). This set of genes, listed in Supplementary Table 6, included CitPH1 (LOC18037376, magnesium-transporting ATPase, P-type 1), two existing versions of CitPH5; namely, CitPH5.2 (LOC18035739, ATPase 10, plasma membrane-type) and CitPH5.1 (LOC18035736, ATPase 10, plasma membrane-type), CitAN1 (LOC18047507, basic helix-loop-helix protein A, synonymous Noemi), CitPH3 (LOC18038669, WRKY transcription factor 44), CitERF13 (LOC18047942, ethylene-responsive transcription factor 13), CitAN11 (LOC18032473, protein TRANSPARENT TESTA GLABRA 1), CitSO (LOC18039929, protein PIN-LIKES 6), CitVHA-c4 (LOC18041768, V-type proton ATPase 16 kDa proteolipid subunit), CitMAC9F1 (LOC18037289, uncharacterized LOC180372899), and CitPH4 (LOC18053295, transcription factor MYB34) (Shi et al., 2015; Li et al., 2016; Shi et al. 2019; Butelli et al., 2019; Strazzer et al. 2019; Huang et al., 2021, Wang et al., 2021). Although most of these genes showed relatively low TPM (Transcripts per Million) values (Figure 2), generally combined with high variability between replicates, raw data show that 8 out of these 11 transcripts, namely, CitPH1, CitPH5.2, CitMAC9F1, CitAN1, CitPH3, CitPH4, CitPH5.1, and CitERF13, were expressed at lower levels in the palatable mandarins. Further qPCR analyses confirmed these tendencies in the samples tested (Supplementary Table 7).

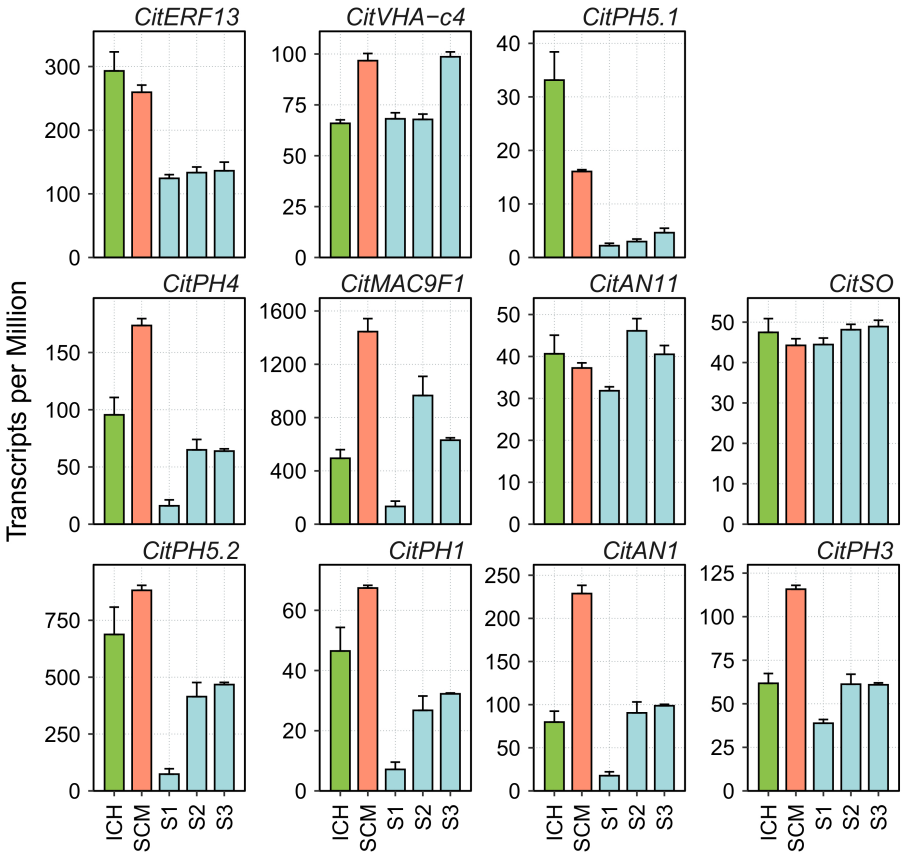


Figure 2. RNA abundance of relevant genes reported to be involved in acid regulation of citrus fruits (Aprile et al., 2011; Butelli et al., 2019; Strazzer et al., 2019; Huang et al., 2021), obtained in RNA-seq analyses of the pulp of developing fruitlets of ICH, SCM, and S1, S2, and S3 segregants. Transcripts per Million were computed using DESeq2 read count normalization. Vertical bars represent standard error from three biological replicates.

Correlation between expression of these genes and acidity, was studied in mature fruit, since the period of acid accumulation in mandarins usually starts about early July and reaches maximum acid levels around the end of September (Cercós et al., 2006). During ripening, fruits of SCM and the three palatable mandarins contained approximately the same sugar quantities (°Br), although SCM fruits were much more acidic, which resulted in lower, unacceptable maturity indices (Supplementary Figure 7). Total acidity in the segregants reached palatability levels similar to those of commercial Clementine and W. Murcott, while SCM still contained higher and unpleasant amounts of acids at the end of the ripening period. qPCR data from juice vesicles of samples collected in November showed that transcript levels of CitVHA-c4, CitPH1, CitPH5.2, CitAN1, CitPH3, CitSO and CitERF13 were downregulated in the three segregants in comparison with SCM, as observed in Clementine and W. Murcott fruits (Figure 3).

Taken together, these observations suggest that CitPH5.1, CitPH4, CitMAC9F1, CitAN11 appear to play a minor role controlling acidity during ripening.

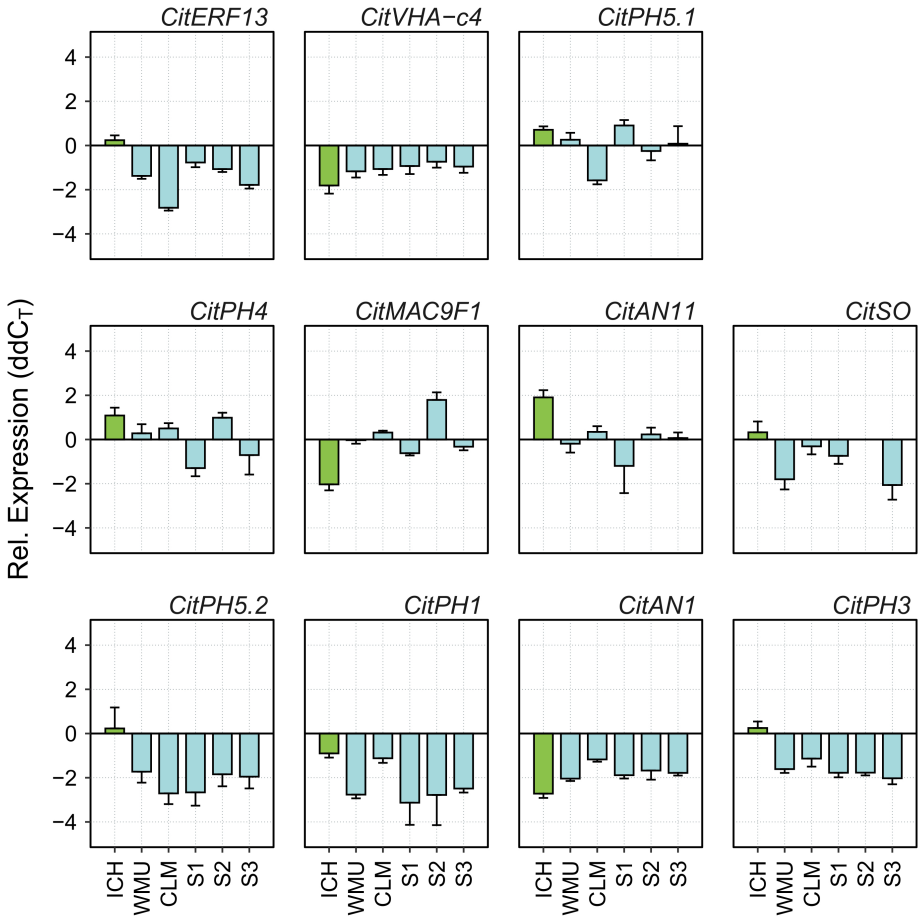


Figure 3. Relative expression of genes reported to be involved in acid regulation of citrus fruits (Aprile et al., 2011; Butelli et al., 2019; Strazzer et al., 2019; Huang et al., 2021), determined by RT-qPCR in juice vesicles of ripening fruits (November) of ICH, SCM, WMU, CLM and S1, S2, and S3 segregants. Vertical bars represent standard error from two technical replicates.

Differential Allele Expression in Pummelo/Mandarin Genetic Admixtures

We additionally studied the influence of pummelo alleles on the differential gene expression of the three segregants as related to SCM. The data show that the percentage of genes expressed in each haplotype sequence, i.e., MA/MA, PU/MA and PU/PU, in the three segregants was, on average, 73%, 25% and 1%, respectively. However, the frequency of DEGs calculated for each haplotype was higher in PU/PU (0.33), intermediate in PU/MA (0.16) and lower in MA/MA (0.09) sequences (Supplementary

Figure 8). As related to the differential expression of the pummelo and the mandarin alleles, in non-DEGs, the percentage of expressed alleles was slightly higher for MA (52-58%) than for PU (42-48%), while in the set of DEGs the PU allele was predominantly expressed (68-75%), in detriment of the MA allele (25-32%). There also was a clear-cut difference between both alleles when the expression trend is considered since most MA (8 out of 12) alleles were downregulated, whereas virtually all PU alleles (27 out of 28) were upregulated. In the three segregants, 4 genes located in pummelo introgressed areas (LOC18035377, tropinone reductase homolog At2g29170; LOC18042174, pectinesterase/pectinesterase inhibitor, PPE8B; LOC18045400, auxin response factor 4; LOC112098377, disease resistance RPP8-like protein 3), only exhibited expression (down or up) of the MA alleles (Figure 4). In contrast, the MA alleles of other 9 genes (LOC18040371, sugar transporter ERD6-like 18; LOC18042131, UDP-glycosyltransferase 83A1; LOC112096160, tropinone reductase homolog At2g29170-like; LOC18043735, potassium transporter 5; LOC112095422, probable pectinesterase/pectinesterase inhibitor 21; LOC18041614, protein DDB_G0271606; LOC18041176, 3-oxo-Delta(4,5)-steroid 5-beta-reductase; LOC1804023, probable linoleate 9S-lipoxygenase 5; LOC18040524, probable pectinesterase /pectinesterase inhibitor 25) were not expressed at all. These set of genes expressing only PU alleles were all upregulated.

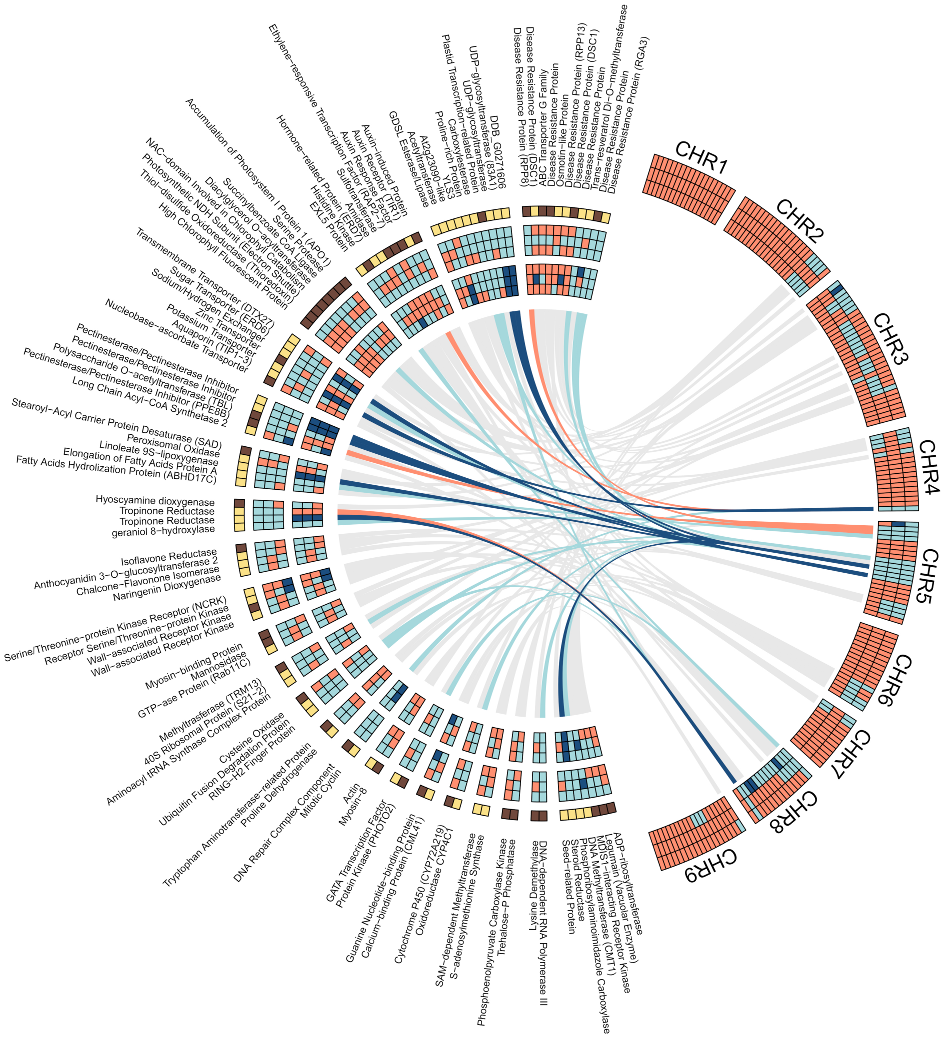


Figure 4. Shared DEGs in the pulp of developing fruitlets of segregants, S1, S2 and S3, as related to SCM. Right Circos. Chromosome admixture structure of the three segregants. From top to bottom, S3 (first row), S2 (second row), and S1 (third row); orange: mandarin/mandarin; light blue: pummelo/mandarin; dark blue: pummelo/pummelo. Left Circos. First row (from top to bottom): gene expression; yellow: up; dark brown: down. Second, third and fourth rows: gene genotype; orange: mandarin/mandarin; light blue: pummelo/mandarin; dark blue: pummelo/pummelo. Fifth, six, and seventh rows: allele expression; orange: mandarin/mandarin; light blue: pummelo/mandarin; dark blue: pummelo/pummelo. Ribbons connect each gene which its physical position in the reference genome. DEGs with the same genotype in the three segregants are connected by colored ribbons; orange: mandarin/mandarin; light blue: pummelo/mandarin; dark blue: pummelo/pummelo. All other DEGs are connected to their respective positions by grey ribbons. DEGs at the outside of the Circos are clustered by categories (Supplementary Table 5) and ordered by the number of members of the category, from top to bottom: Plant defense (11); Uncharacterized process (10); Hormonal regulation (9); Chloroplast (8); Transport (7); Cell wall (5); Lipid and fatty acids (5); Alkaloids (4); Flavonoids (4); Receptors, kinases, transduction (4); Trafficking, vesicles (3); Translation (3); Ubiquitination (3); Amino acid metabolism

(2); Cell division (2); Cytoeskeleton (2); Light signaling (2); Membranes, RE, signaling (2); Redox (2); S-adenosylmethionine metabolism (2); Sugar metabolism (2); and Transcription (2). A final cluster included those categories with a single member: Seed, embryo development (1); Organ development, differentiation (1); Nucleobases (1); Nucleic acids processing (1); Gametophyte and fertilization (1); Cell death (1); and Abiotic stress (1). Results were plotted in R language and environment (R Core Team, 2021), using packages included in Tidyverse collection (Wickham et al., 2019) and circlize (Gu et al., 2014).

Discussion

The main goal of this work was to identify major domestication traits in citrus, based on the comparisons of gene expression patterns in the pulp of developing fruitlets of inedible and edible citrus types. The citrus examined were wild inedible Ichang papeda (ICH; *C. ichangensis* Swingle), acidic Sun Chu Sha Kat mandarin (SCM; *C. reticulata*, Blanco; *C. erythrosa* Tanaka) and three selected palatable genetic admixtures, S1, S2 and S3, derived from a cross between Clementine (CLM; *C. x clementina* Hort. ex Tanaka) and W. Murcott (WMU, *C. x reticulata* Blanco). According to Wang et al., (2017), there were two main mandarin domestication events that generated two mandarin subpopulations differentiated by the degree of acidy. The parentals CLM and WMU were selected for this study because they are representative commercial mandarins of the two adjacent clades of the low acidity subpopulation of mandarins. We carried out the comparison with the three segregants rather than with the parental varieties, to reduce the number of false positives that could be generated comparing two genetically related varieties. The three segregants were selected for the study because their fruits exhibited morphological parameters (Figure 5) and organoleptic traits (Supplementary Figure 7) in the range shown by the parent varieties. ICH, that grows in a truly wild state, is an endemic citrus thought to be originated in glacial refugia in Wuling Mountains and Ta-pa Mountains in southwestern and middle-west China (Yang et al., 2017). It is currently found in natural populations in these areas, is the most cold-resistant citrus and is also tolerant to both damp and drought conditions (Swingle and Reece, 1967; Yang et al., 2017). ICH is considered one of the most primitive wild forms of citrus, produces inedible fruits with very little flesh and juice, if any, and contains acrid and sour oils that release aroma reminiscent of lemons. Recent developments suggest that ICH split from the main citrus clade around 7 million years ago (Wu et al., 2018). According to Tanaka (1954), SCM is an antique mandarin that was very common in temperate China, occurred in Assam and was also cultivated in Japan. The

small SCM fruits, as those of several other traditional mandarins (Wu et al., 2018), are acidic or acidic-sweet, moderately sharp to the taste and very spicy. The ancestor of SCM probably appeared during the last 1.4 million years, after the divergence of the two main subspecies of mainland Asian mandarins (Wu et al., 2021). Under a genomic point of view, ICH and SCM contain pure genomes, i.e., do not show foreign genome introgressions (Wu et al., 2018), while S1, S2, and S3, are genetic admixtures carrying pummelo introgressions, in a mandarin genome background (Wu et al., 2014).

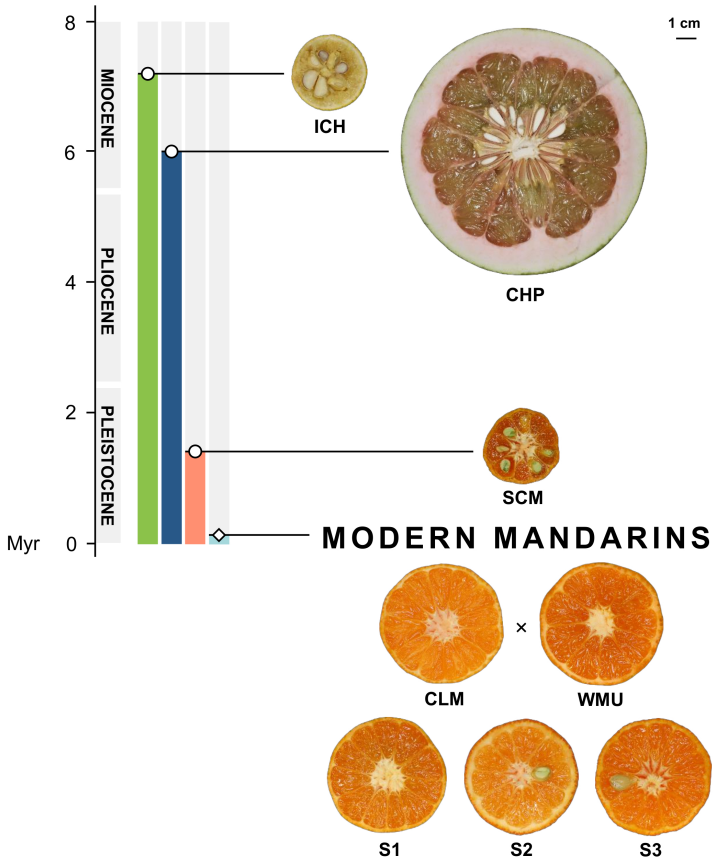


Figure 5. Timeline showing the emergence of the citrus mentioned in this study, according to Wu et al. (2018; 2021). ICH = ichang papeda (*C. ichangensis*), CHP = Chandler pummelo (*C. maxima*), SCM = Sun Chu Sha Kat mandarin (*C. reticulata*), CLM = Clementine mandarin (*C. clementina*), WMU = W. Murcott mandarin (*C. reticulata*), and segregants S1, S2 and S3. The hybrid origin of these three segregants is marked with a cross.

Samples used for RNA-seq analyses were collected during the transition between the phases of cell division and cell elongation of citrus fruits (Cercós et al., 2006; Tadeo et al., 2020), a period that appears to be critical for the establishment of major ripening characteristics (Terol et al., 2019). A first comparison between ICH and SCM rendered

7267 DEGs (Supplementary Table 2), that were mapped (Supplementary Figures 4.001-4.117) against the pathway collection of the KEGG database (Kanehisa et al., 2016). The remaining genes were filtered using more stringent criteria and a final set of 2832 genes were ranked according to the Uniprot database annotation (Uniprot Consortium, 2021) in 45 categories (Supplementary Table 3). In a second analysis, the expression of the three palatable segregants, S1, S2 and S3, was studied as related to SCM. This analysis identified 357 DEGs that showed similar expression in the three palatable mandarins but opposite expression in SCM (Supplementary Table 5). The discussion that follows below highlights the most relevant results derived from those transcriptomic comparisons, while the rest of results are discussed in Supplementary Material.

Differential Gene Expression in Ichang Papeda versus Sun Chu Sha Kat Mandarin: Upregulation of Secondary Metabolism

In ICH, expression of pivotal genes controlling secondary metabolites, as illustrated in the categories Alkaloids, Terpenoids, Phenylpropanoids, Flavonoids, Glucosinolates and Cyanogenic glucosides (Figure 1), was typically upregulated. Regarding Alkaloids, the data indicate that there are several pathways that were promoted in the papeda. Thus, the synthesis of caffeine, a methylxanthin which acts as a natural defense compound (Nathanson, 1984), appears to be promoted, as is that of morphine (Supplementary Table 3) in the isoquinoline alkaloid pathway. In this route, however, the data also suggest that the syntheses of both dopamine and tyramine and also of serotonin, an indoleamine, are downregulated. Interestingly, it has been recently shown that suppression of serotonin biosynthesis increases resistance to insect pests (Lu et al., 2018). Several biosynthetic genes controlling the conversions that render the iridoid glycoside secologanin, the building unit in the biosynthesis of indole and isoquinoline alkaloids, were upregulated in the monoterpenoid pathway (Supplementary Table 3). Iridoid glycosides show a broad defensive spectrum due to their deterrent character on herbivorous and the post-ingestive toxic effects on fungal pathogens (Biere et al., 2004). The upregulation of the synthesis of secologanin and that of other precursors of the indol alkaloids, such as ajmaline, vinblastine and vincristine, may also indicate that the indole alkaloid pathway, was similarly upregulated in the papeda.

This species also showed upregulation of practically all biosynthetic genes of the mevalonate pathway, leading to the isoprenoid intermediates, isopentenyl-PP and dimethylallyl pyrophosphate-PP. The data suggest that two genes encoding geranylgeranyl-PP, the enzyme responsible for the synthesis of the precursors of the main terpenoid groups, were also upregulated (Supplementary Figure 4.072). The synthesis of mono-, sesqui-, and homoterpenoids, the crucial groups of terpenoid volatiles operating attracting parasitoids or repelling herbivores, were characterized in the papeda by a relative high activity. Thus, most of the genes involved in the biosynthesis of the monoterpene limonene, a natural insecticide, antifeedant, antifungal and attractant for pollinators (Erasto and Viljoen, 2008), were upregulated, as it was the synthesis of the phytoalexin, solavetivone, a potent antifungal sesquiterpenoid. In this group, transcripts of α-farnesene synthase (Supplementary Figure 4.078), generating farnesene, that acts as pheromone, a natural insect repellent, and α-copaene synthase-like, that renders several bicyclic olefins and sesquiterpene hydrocarbons, were in contrast expressed at lower levels. There were also two (3S,6E)-nerolidol synthase 1 genes expressed in opposite directions (Supplementary Figure 4.078, Supplementary Table 2). This enzyme participates in the synthesis of the homoterpene 4,8-dimethyl-1,3,7-nonatriene, DMNT, (Degenhardt and Gershenzon, 2000), an observation linked to the upregulation of trimethyltridecatetraene/dimethylnonatriene synthase, encoding the enzyme (EC: 1.14.14.58), generating 4,8,12-trimethyl-1,3(E),7(E),11-tridecatetraene, TMTT, in the Diterpenoid Pathway (Supplementary Figure 4.074). DMNT and TMTT are two irregular acyclic homomoterpenes exhibiting pivotal roles attracting parasitoids and predators of herbivores (Tholl et al., 2011). The synthesis of oryzalexin, a diterpenoid phytoalexin, also appears to be upregulated in the papeda (Supplementary Table 3, Schmelz et al., 2014). In the triterpenoid group, the synthesis of β-amyrin (Supplementary Figure 4.078, Supplementary Table 3), a common plant saponin with important antimicrobial, antifungal, and anti-feedant properties (Faizal and Geelen, 2013), was similarly upregulated. Regarding carotenoids, only transcripts coding for enzymes mediating phytoene synthesis were upregulated in the papeda. Mandarins exhibit a wide range of carotenoids (Figure 5, Gross, 1987), and accordingly showed upregulation of this pathway (Supplementary Figure 4.076). In the phenylpropanoid pathway, main steps in the synthesis of the flavonoid precursors are downregulated, although the synthesis of flavonols, anthocyanidins and anthocyanins were clearly upregulated in the papeda (Supplementary Figures 4.082 and 4.083, Supplementary

Figure 4). Citrus fruits contain a wide range of flavonoids (Nogata et al., 2006), although mandarins and most cultivated citrus species do not, because carry defectives alleles of Ruby gene encoding a MYB transcription factor controlling anthocyanin biosynthesis (Butelli et al., 2017; Wu et al., 2021). From data in Supplementary Figure 4.082, it is suggested that Ruby (synonymous AN2) may participate in the regulation of the expression regulation of naringenin 3-dioxygenase, flavonol synthase and anthocyanidin synthase, key players in the synthesis of anthocyanidins. Flavonoids show antipathogenic activity and participate in the defense against biotic stresses caused by herbivory and pathogenicity. For instance, many flavonoids including the flavonols, kaempferol, quercetin and myricetin may act as deterrents against insects. Flavonoids also reduce the effects of abiotic stresses, such as UV radiation and heat, and show relevant antioxidant properties (Mierziak et al., 2014). In the pathway of the non-flavonoid polyphenol, trans-resveratrol di-O-methyltransferase, the last step in the biosynthesis of the antifungal phytoalexin pterostilbene, is expressed at higher levels in the papeda (Supplementary Table 3). The data also suggest that the synthesis of glucosinolates and cyanogenic glucosides, two kinds of phytoanticipins, may be promoted in the papeda. These are constitutive chemicals, whose non-toxic forms and the catalyzing enzymes that release the toxic compounds are stored in different cells (glucosinolates) or subcellular compartments (cyanogenic glucosides) (Yactayo-Chang et al., 2020).

The depletion of defensive chemicals is a process generally linked to domestication (Moreira et al., 2018), and in SCM appears to have played a critical role in the production of tastier and more flavorful citrus, since these compounds are essentially of bitter taste and toxic to arthropods and vertebrates (Matsuura and Fett-Neto, 2015). It is interesting also to mention that while chemical defenses, secondary metabolites that represent a major barrier to herbivorous insects, are restricted in SCM, in the Plant defense category (Supplementary Material), populated by all kind of protein-based defenses against microbial pathogens, there are more genes up than downregulated.

Differential Gene Expression in Ichang Papeda versus Sun Chu Sha Kat Mandarin: Downregulation of Growth

Another important difference in gene expression between both species is the prevalence in the papeda of downregulation of many genes (>60%), involved in practically all

processes of growth and development (Supplementary Table 3). Categories enriched with downregulated genes included genes with roles in Development, Chloroplast, Hormonal Regulation, Signalling, Gene Expression, and Cellular Growth (Figure 1). While it may be reasonable to find expression of genes involved in chloroplast metabolism or even photosynthesis in the developing fruit pulp, where the transition chloroplast/chromoplast maybe still ongoing, the finding that a high number of genes unquestionably involved in processes such as flowering, fertilization, or organ development (leaves, gametophytes, roots, etc.) were also expressed in this fruit tissue, might be unexpected. There are no convincing explanations for this observation, except perhaps that certain transcripts have not been yet degraded, or that in addition to the reported functions, some genes could also be implicated in fruit growth in hitherto unknown roles. Repressed genes controlling papeda development were also associated with the biosynthesis of cytokinins, brassionosteroids, ethylene (Supplementary Figures 4.077, 4.075 and 4.024), or polyamines (Supplementary Table 4, Supplementary Figure 4.029) and with the transduction of plant hormones in general (Supplementary Figure 4.109). In addition, Supplementary Table 3 also reports that a majority of DEGs included in the Hormonal Regulation category, were downregulated in ICH, particularly those linked to auxins and gibberellins. The repression is also evident in groups of genes related to cellular growth (Figure 1), including the categories of Cell Division (Supplementary Figures 4.099 and 4.102-4.105), Meiosis, Cytoskeleton and Cell Wall (Supplementary Table 3). Consistently, the data also revealed lower levels of gene expression in basic genetic processes regulating growth, such as nucleocytoplasmic transport (Supplementary Figure 4.094), mRNA surveillance pathway (Supplementary Figure 4.095), the processing of the nucleic acids or the chromatin condensation and transcription (Supplementary Table 3). Similarly, low expression was associated with signaling transduction pathways involving genes in Receptors, Kinases, Transduction category or GTPases and second messengers as reported in Supplementary Table 3. Biosynthesis of steroids, one of the most important components of the cellular membranes was also strongly repressed in the papeda (Supplementary Figure 4.012). Other categories enriched with downregulated genes were Transport and Plant Defense (Supplementary Table 3).

Differential Gene Expression in Ichang Papeda versus Sun Chu Sha Kat Mandarin: Activation of Secondary Metabolism versus Growth Stimulation

The RNA-seq analysis, overall, reveals that upregulation in the papeda was mostly associated with the increase of chemical defenses (Table 1), a situation that may imply a penalty in terms of energy and development, as suggested by the downregulation of relevant DEGs involved in a wide variety of growth processes. These conspicuous differences appear to be related to the control of the carbon flux through central pathways of the primary metabolism. Thus, the KEGG data show that gene expression of practically all genes encoding enzymatic activities involved in glycolysis (Supplementary Figure 4.001), cytoplasmatic citric acid degradation (Supplementary Figures 4.058, 4.017, and 4.021), GABA shunt (Supplementary Table 2), fatty acid degradation (Supplementary Figure 4.010), TCA cycle (Supplementary Figure 4.002), and several subunits of the major regulatory complex of the oxidative phosphorylation process (Supplementary Figure 4.014) were upregulated in the papeda. As in SCM, in the highly acidic species lemon and citron, several genes involved in the TCA cycle and GABA shunt also displayed reduced expression during ripening (Borredá et al., 2022). Based on these observations, we propose that in the papeda, the generation of energy in the ATP form, is stimulated through the increase of the carbon flux via both glycolysis and fatty acid degradation, generating pyruvate and acetyl CoA, respectively, to fuel the Krebs cycle (Figure 6). The activation of the TCA cycle increases the production of both succinate, a substrate of complex II of the mitochondrial electron transport chain, and citric acid, that after transport to the cytosol may increase cytoplasmatic acidity to a detrimental level for normal cellular functions. Citric acid may be, then, stored in the vacuole, catalyzed to Acetyl CoA or further metabolized into glutamate entering the GABA shunt, that finally restores the carbon pool that the TAC cycle requires. We proposed in a previous work (Cercós et al., 2006), that in cultivated Clementine the GABA shunt, a powerful proton consuming reaction, is a very efficiently way to reduce both citric acid and cytoplasmatic acidity in ripe fruit flesh, while this current work suggests that this mechanism is active in developing fruitlets of wild citrus. Current consensus agrees that the regulation of acid metabolism in citrus, is basically focused on the generation of citric acid in the TCA, and its storage in the vacuole and later reduction in the cytosol through the GABA and ATP citrate lyase pathways (Cercós et

al., 2006; Sadka et al., 2019; Li et al., 2019, Feng et al., 2021). However, our suggestion expands this concept identifying the regulatory TCA as a hub linking catabolism of fatty acids, production of organic acids and activation of oxidative phosphorylation, for the generation of energy as requested by growth and/or environmental demands. In this view, citric acid appears to be the major player in a system, that is balanced modulating its concentration and compartmentalization through processes that ultimately determine fruit acidity, a pivotal citrus organoleptic trait.

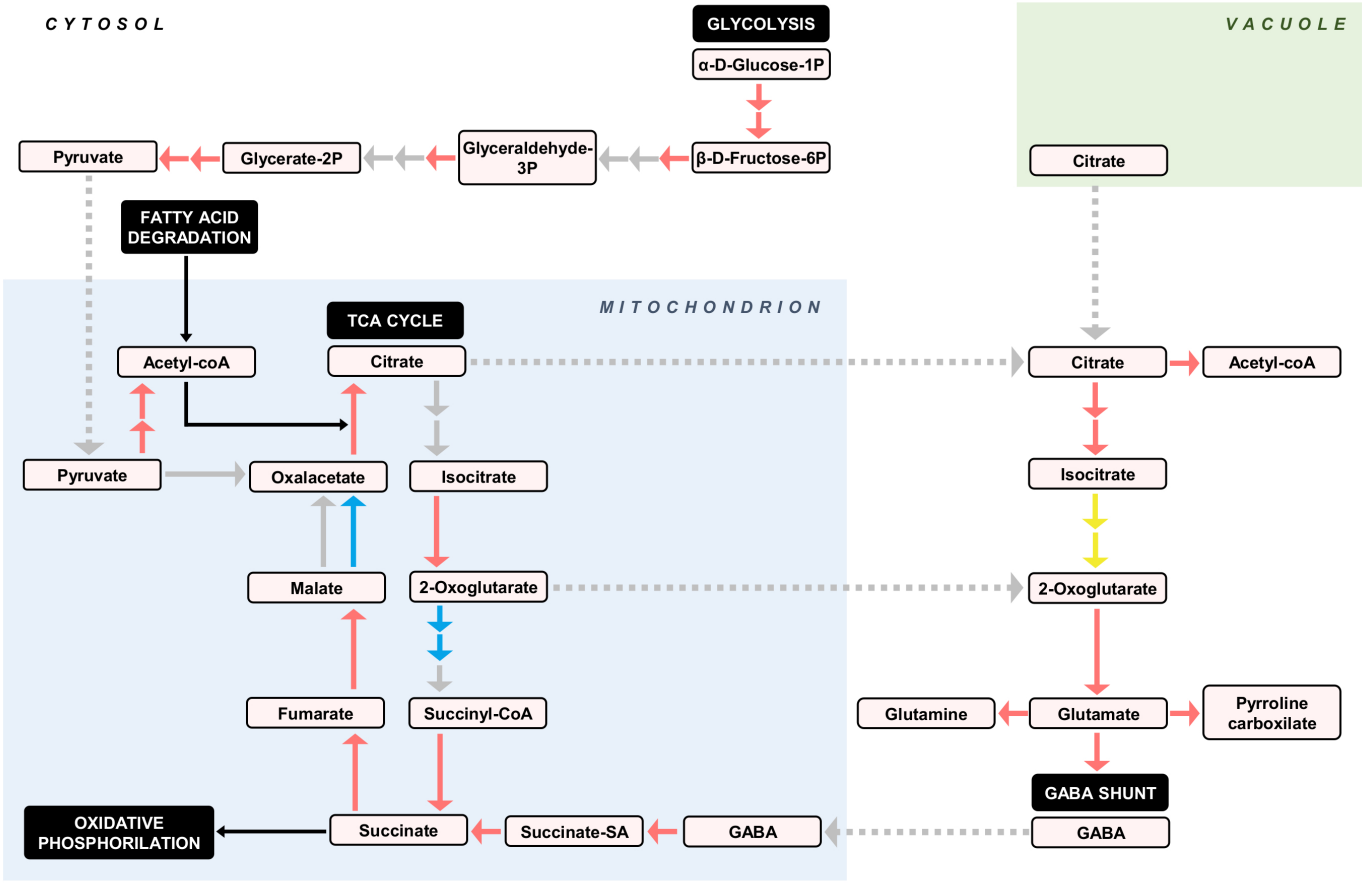


Figure 6. Proposed activity of central carbon metabolic pathways, deduced from DEGs in the pulp of developing fruitlets of ICH as related to SCM. Bigger solid arrows indicate gene regulation: red = upregulation, blue = downregulation, yellow = undetermined, and grey = no differential expression. Small solid black arrows represent substrate inputs, while grey dotted arrows indicate directional transport. Reaction products and pivotal pathways are embedded in colored and black boxes, respectively.

These changes differentially affected nitrogen and carbon allocation in both species. The most important difference in amino acid production was probably related to the degradation of leucine, isolecucine and valine, that was promoted in the papeda (Supplementary Table 4). These amino acids are degraded to Acetyl CoA and succinyl CoA that may thus fuel the TCA (Supplementary Figures 4.025, and 4.002). In addition, upregulation in the papeda, of practically all genes controlling fatty acid degradation (Supplementary Figure 4.010), also appears to contribute to provide higher amounts of Acetyl-CoA to fuel the citrate cycle. In addition, the degradation of leucine produces hydroxymethylglutaryl, an intermediate of the mevalonate pathway in the terpenoid pathway, that may then be reinforced (Supplementary Figure 4.072). This pathway also provides precursors for monoterpenoid and isoquinoline alkaloid (Supplementary Figure 4.087), whose synthesis appears to be upregulated in the papeda as suggested above. Synthesis of major intermediates participating in carbon primary metabolism, for instance, was apparently more active in ICH (Supplementary Figure 4.045). In contrast, the number of upregulated DEGs involved in transport of hexoses and monosaccharides, including the vacuolar hexose transporter SWEET17 and NDR1/HIN1-like protein 26, required for correct sugar partitioning between source leaves and sink organs, was higher in SCM mandarin than in ICH (Supplementary Table 3).

The above results, overall, indicate that growth and development is rather restricted in inedible wild ICH, while secondary metabolism and the production of chemical defenses in particular (Table 1), are clearly upregulated. The dichotomy between growth stimulation versus activation of secondary metabolism is a common situation in the plant kingdom, which poses to plants, as sessile organisms, a dilemma that is resolved balancing the cost of investment in chemical defense and the availability of resources for its development. The payment of these costs, which takes place in the form of energy and carbon and nitrogen supplies, implies a proportional reduction in the growth and development of the plant (Mithöfer and Boland, 2012). It has been indicated, for instance, that pathogen and insect tolerance and resistance of domesticated citrus has generally declined compared with wild relatives (Bernet et al., 2005). We have previously shown through genomic analysis that citrus domestication tended to reduce chemical defenses involving cyanogenesis and alkaloids (Gonzalez-Ibeas et al., 2021b), while in the current work we expand this concept and show evidence that practically all

major groups of chemical defenses, including alkaloids, terpenoids, glucosinolates, and cyanogenic glycosides are repressed in SCM. Therefore, the results support the suggestion that in our system, the papeda restricts its growth to allocate resources and energy to the production of defensive chemicals to escape herbivory.

Table 1. Expression of genes involved in the biosynthesis of chemical defenses and associated compounds, in the pulp of developing fruitlets of ICH as related to SCM.

Chemical Defense	Gene Id	Gene Name	Compound	Expression
Alkaloids	LOC18037906	vinorine synthase	Ajmaline	Up
Alkaloids	LOC112100740	probable caffeine synthase 4	Caffeine	Up
Alkaloids	LOC18036840	codeine O-demethylase	Morphine	Up
Alkaloids	LOC18037214	8-hydroxygeraniol dehydrogenase	Vinblastine	Up
Alkaloids	LOC18042369	7-deoxyloganetin glucosyltransferase	Secologanin	Up
Alkaloids	LOC18043348; LOC18046227	tyrosine/DOPA decarboxylase 5	Serotonin	Down
Alkaloids	LOC18041681	tryptamine 5-hydroxylase (CYP71P1)	Serotonin	Down
Alkaloids	LOC18043348; LOC18046227	tyrosine/DOPA decarboxylase 5	Dopamine	Down
Terpenoids	LOC112098571;	(R)-limonene synthase 1, chloroplastic-like	Limonene	Up

	LOC112098486			
Terpenoids	LOC18045727	beta-amyrin synthase	ß-amyrin	Up
Terpenoids	LOC18049179; LOC18049178	cytochrome P450 82G1	TMTT	Up
Terpenoids	LOC18051782	(3S,6E)-nerolidol synthase 1	DMNT	Up
Terpenoids	LOC18033168	(3S,6E)-nerolidol synthase 1	DMNT	Down
Terpenoids	LOC112095677	α-copaene synthase-like	α-copaene, (E)-β-caryophyllene and germacrene D	Down
Terpenoids	LOC18053589	α-farnesene synthase	Farnesene	Down
Phytoalexins	LOC18031421	cytochrome P450 76M5	Oryzalexin	Up
Phytoalexins	LOC18052043	premnaspirodiene oxygenase	Solavetivone	Up
Phytoalexins	LOC18031586; LOC18051743	trans-resveratrol di-O-methyltransferase	Pterostilbene	Up

Flavonoids	LOC18037475	flavonol synthase/flavanone 3-hydroxylase	Kaempferol	Up
Flavonoids	LOC18037475	flavonol synthase/flavanone 3-hydroxylase	Quercetin	Up
Flavonoids	LOC18037476	flavonol synthase/flavanone 3-hydroxylase	Myricetin	Up
Cyanogenic glucosides	LOC18036876	linamarin synthase 1	Linamarin	Up
Glucosinolate metabolism	LOC18049889	flavin-containing monooxygenase FMO GS-OX-like 4	Methylsulfinylalkyl glucosinolates	Up

Differential Gene Expression in Ichang Papeda versus Sun Chu Sha Kat Mandarin: Upregulation of Cold Tolerance

The RNA-seq analysis, on the other hand, did not provide strong indications or evidence that both species behave differently facing abiotic stresses (Figure 1), except for several genes involved in cold and oxidative stresses (Supplementary Table 3). The set of DEGs related to cold was mostly characterized by the upregulation of genes implicated in cellular responses, including CORA, a cold and drought-regulated protein that showed the highest expression observed in this analysis (Jha et al., 2021). Key transcription factors governing cold response genes were also differentially expressed between both species. The most striking difference was related to the absence in the papeda of MYBS3 mRNA, a central transcription repressor that suppresses the DREB1/CBF-dependent signaling pathway regulating cold stress responses (Su et al., 2010). Also noticeable was the downregulation of negative regulators of abscisic acid such as MSI4 (Banerjee et al., 2017) and ERD15 (Kariola et al., 2006), components of stress responses, including freezing resistance. Furthermore, JUB1, a gene that modulates cellular H_2O_2 levels (Wu et al., 2012), enhancing tolerance to various abiotic stresses including cold (Fang et al., 2021), was upregulated (Supplementary Table 3). It is also worth to highlight that DEGs related to sphingolipid metabolism were repressed, except the conversion of ceramide to phytoceramide-1-phosphate (Supplementary Figure 4.054), that appears to be important for the resistance to cold (Dutilleul et al., 2015). Taken together, these observations might be related to the fact that ichang papeda is the hardiest species in the genus *Citrus* (Swingle and Reece, 1967), tolerating both frost temperatures, even at -20°C, and damp conditions (Yang et al. 2017). It should be notice that these observations were made in samples not subjected to cold conditions, while the natural habitat of the papeda in montane regions of China (Yang et al. 2017) is rather chiller.

Since temperature stress increases the generation of ROS (Hasanuzzaman et al., 2013), this circumstance might be also connected with the enrichment of upregulated DEGs with several roles on antioxidant defense, that were detected in the Redox category (Supplementary Table 3), in the salvage pathway of the central coenzyme NAD^+/NADH (Supplementary Figure 4.067) and in the synthesis of other coenzymes and vitamins, such as pantothenate (vitamin B5) and CoA (Supplementary Figure 4.068), riboflavin and flavin mononucleotide (Supplementary Figure 4.065), tocopherol and tocotrienol

(vitamin E) (Supplementary Figure 4.013), and the compounds integrating vitamin B6 (Supplementary Figure 4.066).

Differential Gene Expression in Mandarin Admixtures versus Sun Chu Sha Kat Mandarin: Palatability Increment

The comparative RNA-seq analyses between the four mandarin transcriptomes was characterized by the predominant downregulation in the three segregants of genes involved in both Abiotic Stress and Plant Defense categories, mostly participating in central roles and general responses, such as SRG1, regulating plant immunity. Similarly, pivotal genes implicated in the synthesis of relevant terpenoids, alkaloids and glucosinolates, and hence, in chemical defense, i.e., the pheromone β-farnesene, the homoterpene TMTT, the antifeedant limonene, the alkaloid scopolamine, or benzyl-glucosinolate, were downregulated. It is worth to note that in the wild papeda the expression of cytochrome P450 82G1 and (R)-limonene synthase 1, chloroplastic-like, the regulatory genes controlling the synthesis of TMTT and limonene, was relatively high (Supplementary Table 2 and 3), while in SCM, these genes were expressed at lower levels and in the three palatable mandarins, their expression was hardly detected. However, there were other defense genes (Supplementary Material) upregulated in the three segregants, an effect perhaps related to specific responses to local pathogen attacks and/or to the contribution of pummelo.

The three palatable mandarins showed downregulation of important genes controlling structural components of cell wall such as cellulose, expansins, pectins, and lignans and lignin. Expression of cytochrome P450 84A1 (ferulate 5 hydroxylase) involved in lignin biosynthesis, for example, was high in ICH, lower in SCM (Supplementary Table 2) and even lower in the three segregants (Supplementary Table 5). These observations might suggest that cell wall stiffening of the juice sacs in the fruit pulp or the peel of the segments is reduced, a characteristic that may be associated to a higher degree of palatability.

Other DGEs that may also affect palatability and taste were methanol O-anthraniloyltransferase, and 2-methylene-furan-3-one reductase, that were both upregulated and glutathione S-transferase L3, that was repressed. The first gene is an acyltransferase that catalyzes the formation of methyl anthranilate in the acridone alkaloid pathway, a substance of pleasant aroma, involved in the fragrance of Concord

grapes (Wang and De Luca, 2005), that has been used in flavoring foods as mandarin candies or soft drinks (cited in Luo et al., 2019; Lee et al., 2019). Interestingly, expression of this gene is barely detectable in the inedible papeda, but its expression is relatively high in SCM and even higher in the three segregants. The enone oxidoreductase 2-methylene-furan-3-one reductase, on the other hand, renders furaneol, a key flavor compound in strawberries (Raab et al., 2006), that appear to be present only in fruits. The third gene, glutathione S-transferase L3 catalyzes the reduction of S-glutathionylquercetin to quercetin, a polyphenol that has a bitter flavor, in agreement with the suggestion that in comparison with wild citrus, cultivars show decreased secondary metabolite levels, such as bitterness compounds (Rao et al., 2021).

In citrus, taste is mainly dependent of the sugar and acid content of the juice. Regarding soluble sugars, the three palatable mandarins showed repression of sucrose synthase 2, probable galactinol-sucrose galactosyltransferase 2 and stachyose synthase, suggesting that sucrose conversion and catabolism in the three segregants is limited during this immature stage favoring sucrose accumulation. Citrus fruitlets appear to operate as main utilization sinks and sucrose synthase expression and activity at these immature fruit stages generally are relatively low (Li et al., 2019; Iglesias et al., 2007). During fruit development, upregulating of sucrose synthases enhances sink strength promoting sucrose and starch accumulation and increasing fruit size, (Sadka et al., 2019; Feng et al. 2021). SCM also showed downregulation of two ERD6L sugar transporters that might be relevant for sugar accumulation in citrus fruits, especially sugar transporter ERD6-like 7, as reported in kumquats (Wei et al., 2021).

There were two sodium/hydrogen exchanger genes operating in low affinity electroneutral exchange of protons for cations, expressed in opposite directions in the three segregants. One of these, sodium/hydrogen exchanger 6 is the orthologous of CsNHX, that has been reported to be involved in the regulation of acidity levels in low-acid orange mutants (Wang et al., 2021). In addition, ATPase 10, plasma membrane-type, whose expression has been associated with citric acid accumulation in lemon juice sac cells (Aprile et al., 2011) was upregulated in the acidic mandarin.

Differential Gene Expression in Palatable Mandarin Admixtures versus Sun Chu Sha Kat Mandarin: Downregulation of Acidity

Since acidity is one of the fundamental traits determining palatability of citrus fruits and therefore a critical trait for citrus domestication (Wang et al., 2018; Butelli et al., 2019; Rao et al., 2021) we examined expression of pivotal genes previously suggested or proposed to regulate acidity in citrus (Supplementary Table 6). The analysis of gene expression in fruit pulp revealed that total acidity in palatable mature fruits was linked to downregulation of 5 genes, CitPH1 (magnesium-transporting ATPase, P-type 1), CitPH5.2 (ATPase 10, plasma membrane-type;), CitAN1 (basic helix-loop-helix protein A), CitPH3 (WRKY transcription factor 44) and CitERF13 (ethylene-responsive transcription factor 13), in both developmental stages analyzed, developing and ripening fruits (Figure 3) in all three segregants. Of the genes studied, CitAN11 (protein TRANSPARENT TESTA GLABRA 1) was not repressed in any of the two stages, CitSO (protein PIN-LIKES 6) and CitVHA-c4 (V-type proton ATPase 16 kDa proteolipid subunit) were only downregulated in ripening fruits, whereas CitMAC9F1 (uncharacterized LOC180372899, CitPH4 (transcription factor MYB34) and CitPH5.1, in contrast, were downregulated in developing fruitlets.

The current study, while providing data on mandarins, complements the proposal of Strazzer et al., (2019), that shows that CitPH1 and CitPH5, two major downstream genes involved in vacuolar acidification, are highly expressed in ripe fruits of acidic varieties of lemons, oranges, and pummelos. Expression of both CitPH5 (Shi et al., 2015; Feng et al., 2021; Shi et al., 2019) and AN1 (Butelli et al., 2019, Strazzer et al., 2019, Wang et al., 2021), has been associated with citric acid accumulation in a number of studies, a subject revised in Huang et al., (2021).

CitPH1 and CitPH5 appear to act together since in petunia, PH1 may bind to PH5 to promote PH5 proton-pumping activity (Faraco et al., 2014). CitPH1 and CitPH5 expression, in contrast, is strongly reduced in acidless varieties of citrus and this downregulation is associated with mutations that disrupt expression of CitPH4 (MYB), CitAN1 (HLH) and/or CitPH3 (WRKY) transcription factors (Strazzer et al., 2019). These authors also report that CitMAC9F1, a gene of unknown function, is activated by the same transcription factors as CitPH1 and CitPH5 and that CitSO does not contribute to the differences in acidity. Our analyses are in line with these results since in both

developing and ripening mandarin fruits CitPH1, CitPH5.2, CitAN1, and CitPH3 were downregulated. In addition, developing fruits showed repression of CitMAC9F1, CitPH4, and CitPH5.1, while ripe fruits exhibited further downregulation of CiSO (Figures 2 and 3).

It has also been proposed that CitERF13 regulates citrate accumulation by directly activating the vacuolar proton pump gene CitVHA-c4102 (Li et al., 2016). In mandarins, CitERF13 was effectively downregulated in young and mature fruits while CitVHA-c4, that appears to be normally expressed during ripening (Feng et al., 2021) was repressed only in young fruit. The above observations suggest that although this set of genes, except CitAN11, is very likely involved in the control of acidity in the fruit pulp of mandarins, there may be precise patterns of regulation of gene expression, specifically controlling the accumulation and/or degradation of organic acids at each developmental stage.

In the work by Strazzer et al. (2019), the role CitPH1 and CitPH5, was mostly deduced comparing large differences in acidity between acidic and acidless varieties. They reported that those differences are produced by mutations disrupting the expression of those transcription factors that regulate the two ATPases. Consequently, the authors wonder if small acidity differences between varieties of the same group, may also be due to small differences in the expression of CitPH1/CitPH5. The evidence presented in this current work answers this question showing evidence that smaller differences in non-palatable acidic SCM and edible mandarins are equally correlated with the expression of CitPH1 and CitPH5 and that the range of this change is enough to determine its acceptance, suggesting that this circumstance was a pivotal domestication trait in citrus.

Differential Allele Expression in Palatable Mandarin Admixtures

The study on differential allelic expression on the pummelo introgressed areas of the three segregants presented in (Figure 4) indicates that the contribution of pummelo to acidity did not play a critical role. The analysis detected, however, genetic targets that in principle appear to be contributors to other relevance traits, such as sugar transporters, cell wall modifying pectinesterases or auxin responses. For instance, expression of a major regulator of auxin action, the auxin receptor TIR1 (LOC18052162), in the pulp of developing fruitlets is low in the papeda, relatively high in SCM and higher in the three

segregants. Interestingly, differential allelic expression was mostly characterized by the downregulation of MA alleles and the upregulation of the PU ones (Supplementary Figure 8). This fact suggests that mandarin domestication in part was certainly based on the selection and substitution of mandarin alleles by pummelo genes.

In conclusion, we propose that during the transition of inedible papedas to sour mandarins, domestication involved a first phase of major changes in the gene regulation of central pathways of the primary and secondary metabolism, characterized by both growth stimulation and reduction of distasteful chemical defenses. It is intriguing, that this reduction appears to affect to all main alkaloids, except dopamine and serotonin, two brain neurotransmitters regulating mood and emotion in humans. We also suggest that in a second phase, several edible attributes of mandarins, especially acidity, were progressively improved through specific changes. Lastly, several observations might indicate that the strong resilience of Ichang papeda to frost could be related to the regulation of relevant genes governing cold response.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

EP, collected samples and phenotype data, performed data analysis, plot the results, designed figures and tables, and reviewed and edited the original draft. CB, advised data curation. FT, assisted reviewing and editing the original draft. MT, conceived and conceptualized the study, conducted and supervised the research, interpreted the results and wrote and edited the original draft.

Funding

This research is co-funded by the Ministerio de Ciencia, Innovación y Universidades (Spain) through grant #RTI2018-097790-R-100 and by the European Union through the European Regional Development Fund (ERDF) of the Generalitat Valenciana 2014–2020, through grants #PROMETEO/2020/027 and #IVIA/52201A.

Acknowledgements

We thank Matilde Sancho and Antonio Prieto for laboratory tasks and sampling.

Data Availability Statement

The data presented in this study are deposited in the Sequence Read Archive (SRA database) repository, accession number PRJNA853264 (https://www.ncbi.nlm.nih.gov/sra/PRJNA853264).

References

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(4), 551–563. doi: 10.1007/s10142-011-0226-3

Banerjee, A., Wani, S. H., and Roychoudhury, A. (2017). Epigenetic Control of Plant Cold Responses. *Front Plant Sci*, 8, 1643. doi: 10.3389/fpls.2017.01643

Bernet, G. P., Margaix, C., Jacas, J., Carbonell, E. A., and Asins, M. J. (2005). Genetic analysis of citrus leafminer susceptibility. *Theor Appl Genet*, 110(8), 1393–1400. doi: 10.1007/s00122-005-1943-6

Biere, A., Marak, H. B., and van Damme, J. M. (2004). Plant chemical defense against herbivores and pathogens: generalized defense or trade-offs?. *Oecologia*, 140(3), 430–441. doi: 10.1007/s00442-004-1603-6

Borredá, C., Perez-Roman, E., Talon, M., and Terol, J. (2022). Comparative transcriptomics of wild and commercial Citrus during early ripening reveals how domestication shaped fruit gene expression. *BMC Plant Biol* 22, 123. doi: 10.1186/s12870-022-03509-9

Buck, R., and Flores-Rentería, L. (2022). The Syngameon Enigma. *Plants*, 11, 895. doi: 10.3390/plants11070895

Butelli, E., Garcia-Lor, A., Licciardello, C., Las Casas, G., Hill, L., Recupero, G. R., et al. (2017). Changes in Anthocyanin Production during Domestication of Citrus. *Plant Physiology*, 173(4), 2225–2242. doi: 10.1104/pp.16.01701

Butelli, E., Licciardello, C., Ramadugu, C., Durand-Hulak, M., Celant, A., Recupero, G. R., et al. (2019). Noemi controls production of flavonoid pigments and fruit acidity and illustrates the domestication routes of modern citrus varieties. *Current Biology*, 29(1), 158-164. doi: 10.1016/j.cub.2018.11.040

Cercós, M., Soler, G., Iglesias, D. J., Gadea, J., Forment, J., and Talón, M. (2006). Global analysis of gene expression during development and ripening of citrus fruit flesh. A proposed mechanism for citric Acid utilization. *Plant Mol Biol*, 62(4-5), 513–527. doi: 10.1007/s11103-006-9037-7

Chen, J., Li, G., Zhang, H., Yuan, Z., Li, W., Peng, Z., Shi, M., Ding, W., et al. (2021). Primary Bitter Taste of Citrus is Linked to a Functional Allele of the 1,2-Rhamnosyltransferase Gene Originating from Citrus grandis. *J Agric Food Chem*, 69(34), 9869–9882. doi: 10.1021/acs.jafc.1c01211

Deng, X., Yang X., Yamamoto, M., and Biswas M.K. (2020). "Domestication and history" in The Genus Citrus, ed. Talon, M., Caruso, M. and Gmitter F.G., (Woodhead Publishing), 33–55.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. doi: 10.1093/bioinformatics/bts635

Degenhardt, J., and Gershenzon, J. (2000). Demonstration and characterization of (E)-nerolidol synthase from maize: a herbivore-inducible terpene synthase participating in (3E)-4,8-dimethyl-1,3,7-nonatriene biosynthesis. *Planta*, 210(5), 815–822. doi: 10.1007/s004250050684

Dutilleul, C., Chavarria, H., Rézé, N., Sotta, B., Baudouin, E., and Guillas, I. (2015). Evidence for ACD5 ceramide kinase activity involvement in Arabidopsis response to cold stress. *Plant Cell Environ*, 38(12), 2688–2697. doi: 10.1111/pce.12578

Erasto, P., and Viljoen, A. M. (2008). Limonene-a review: biosynthetic, ecological and pharmacological relevance. *Natural Product Communications*, 3(7), 1934578X0800300728

Faizal, A., and Geelen, D. (2013). Saponins and their role in biological processes in plants. *Phytochemistry Rev*, 12(4), 877-893. doi: 10.1007/s11101-013-9322-4

Fang, P., Wang, Y., Wang, M., Wang, F., Chi, C., Zhou, Y. et al. (2021). Crosstalk between Brassinosteroid and Redox Signaling Contributes to the Activation of CBF Expression during Cold Responses in Tomato. *Antioxidants* (Basel), 10(4), 509. doi: 10.3390/antiox10040509

Faraco, M., Spelt, C., Bliek, M., Verweij, W., Hoshino, A., Espen, L., et al. (2014). Hyperacidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color. *Cell Rep*, 6(1), 32–43. doi: 10.1016/j.celrep.2013.12.009

Feng, G., Wu, J., Xu, Y., Lu, L., and Yi, H. (2021). High-spatiotemporal-resolution transcriptomes provide insights into fruit development and ripening in Citrus sinensis. *Plant Biotechnol. J.*, 19(7), 1337–1353. doi: 10.1111/pbi.13549

Gonzalez-Ibeas, D., Ibanez, V., Perez-Roman, E., Borredá, C., Terol, J., and Talon, M. (2021). Shaping the biology of citrus: I. Genomic determinants of evolution. *Plant Genome*, 14(3), e20133. doi: 10.1002/tpg2.20104

Gonzalez-Ibeas, D., Ibanez, V., Perez-Roman, E., Borredá, C., Terol, J., and Talon, M. (2021). Shaping the biology of citrus: II. Genomic determinants of domestication. *Plant Genome*, 14(3), e20133. doi: 10.1002/tpg2.20133

Gross, J. (1987). "Carotenoids" in Pigments in Fruits, ed. Schweigert, B. S., (Academic Press, London), 87–186.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics*, 30(19), 2811–2812. doi: 10.1093/bioinformatics/btu393

Hamston, T. J., de Vere, N., King, R. A., Pellicer, J., Fay, M. F., Cresswell, J. E., et al. (2018). Apomixis and Hybridization Drives Reticulate Evolution and Phyletic Differentiation in Sorbus L.: Implications for Conservation. *Front Plant Sci*, 9, 1796. doi: 10.3389/fpls.2018.01796

Hasanuzzaman, M., Nahar, K., and Fujita, M. (2013). "Extreme Temperature Responses, Oxidative Stress and Antioxidant Defense in Plants" in Abiotic Stress - Plant Responses and Applications in Agriculture, ed. Vahdati, K. and Leslie, C., (IntechOpen), doi: 10.5772/54833

Huang, X. Y., Wang, C. K., Zhao, Y. W., Sun, C. H., and Hu, D. G. (2021). Mechanisms and regulation of organic acid accumulation in plant vacuoles. *Hortic Res*, 8(1), 227. doi: 10.1038/s41438-021-00702-z

Iglesias, D. J., Cercós, M., Colmenero-Flores, J. M., Naranjo, M. A., Ríos, G., Carrera, E., et al. (2007). Physiology of citrus fruiting. *Braz J Plant Phys*, 19(4), 333-362. doi: 10.1590/S1677-04202007000400006

Jha, R. K., Patel, J., Patel, M. K., Mishra, A., and Jha, B. (2021). Introgression of a novel cold and drought regulatory-protein encoding CORA-like gene, SbCDR, induced osmotic tolerance in transgenic tobacco. Physiol Plant, 172(2), 1170–1188. doi: 10.1111/ppl.13280

Jirschitzka, J., Schmidt, G. W., Reichelt, M., Schneider, B., Gershenzon, J., and D'Auria, J. C. (2012). Plant tropane alkaloid biosynthesis evolved independently in the Solanaceae and Erythroxylaceae. *Proc Natl Acad Sci U S A*, 109(26), 10304-10309. doi: 10.1073/pnas.1200473109

Kalita, B., Roy, A., Annamalai, A., and Lakshmi, P. T. V. (2021). A molecular perspective on the taxonomy and journey of Citrus domestication. *Perspectives in Plant Ecology, Evolution and Systematics*, 53, 125644. doi: 10.1016/j.ppees.2021.125644

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, 44(D1), D457–D462. doi: 10.1093/nar/gkv1070

Kariola, T., Brader, G., Helenius, E., Li, J., Heino, P., and Palva, E. T. (2006). EARLY RESPONSIVE TO DEHYDRATION 15, a negative regulator of abscisic acid responses in Arabidopsis. *Plant Physiol*, 142(4), 1559–1573. doi: 10.1104/pp.106.086223

Khoo, H. E., Azlan, A., Tang, S. T., and Lim, S. M. (2017). Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr Res*, 61(1), 1361779. doi: 10.1080/16546628.2017.1361779

Klemens, P. A., Patzke, K., Trentmann, O., Poschet, G., Büttner, M., Schulz, A., et al. (2014). Overexpression of a proton-coupled vacuolar glucose exporter impairs freezing tolerance and seed germination. *New Phytol*, 202(1), 188-197. doi: 10.1111/nph.12642

Köllner, T. G., Held, M., Lenk, C., Hiltpold, I., Turlings, T. C., Gershenzon, J., et al. (2008). A maize (E)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *Plant Cell*, 20(2), 482–494. doi: 10.1105/tpc.107.051672

Lee, H. L., Kim, S. Y., Kim, E. J., Han, D. Y., Kim, B. G., and Ahn, J. H. (2019). Synthesis of Methylated Anthranilate Derivatives Using Engineered Strains of Escherichia coli. *J Microbiol Biotechnol*, 29(6), 839–844. doi: 10.4014/jmb.1904.04022

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN]. https://arxiv.org/abs/1303.3997v2

Li, L. J., Tan, W. S., Li, W. J., Zhu, Y. B., Cheng, Y. S., and Ni, H. (2019). Citrus Taste Modification Potentials by Genetic Engineering. *Int J Mol Sci*, 20(24), 6194. doi: 10.3390/ijms20246194

Li, S. J., Yin, X. R., Xie, X. L., Allan, A. C., Ge, H., Shen, S. L., et al. (2016). The Citrus transcription factor, CitERF13, regulates citric acid accumulation via a protein-protein interaction with the vacuolar proton pump, CitVHA-c4. *Sci Rep*, 6, 20151. doi: 10.1038/srep20151

Liang, M., Cao, Z., Zhu, A., Liu, Y., Tao, M., Yang, H., et al. (2020). Evolution of self-compatibility by a mutant Sm-RNase in citrus. *Nat Plants*, 6(2), 131–142. doi: 10.1038/s41477-020-0597-3

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923–930. doi: 10.1093/bioinformatics/btt656

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12), 550. doi: 10.1186/s13059-014-0550-8

Lu, H. P., Luo, T., Fu, H. W., Wang, L., Tan, Y. Y., Huang, J. Z., et al. (2018). Resistance of rice to insect pests mediated by suppression of serotonin biosynthesis. *Nat Plants*, 4(6), 338–344. doi: 10.1038/s41477-018-0152-7

Luo, W., and Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 29(14), 1830–1831. doi: 10.1093/bioinformatics/btt285

Luo, Z. W., Cho, J. S., and Lee, S. Y. (2019). Microbial production of methyl anthranilate, a grape flavor compound. *Proc Natl Acad Sci U S A*, 116(22), 10749–10756. doi: 10.1073/pnas.1903875116

Matsuura, H. N., and Fett-Neto, A. G. (2015). "Plant Alkaloids: Main Features, Toxicity, and Mechanisms of Action" in Plant Toxins. Toxinology, ed. Gopalakrishnakone, P., Carlini, C. and Ligabue-Braun, R. (Springer, Dordrecht), doi: 10.1007/978-94-007-6728-7_2-1

Merelo, P., Agustí, J., Arbona, V., Costa, M. L., Estornell, L. H., Gómez-Cadenas, A., Coimbra, S., et al. (2017). Cell Wall Remodeling in Abscission Zone Cells

during Ethylene-Promoted Fruit Abscission in Citrus. *Front Plant Sci*, 8, 126. doi: 10.3389/fpls.2017.00126

Mierziak, J., Kostyn, K., and Kulma, A. (2014). Flavonoids as important molecules of plant interactions with the environment. *Molecules*, 19(10), 16240–16265. doi: 10.3390/molecules191016240

Mithöfer, A., and Boland, W. (2012). Plant defense against herbivores: chemical aspects. *Annu Rev Plant Biol*, 63, 431–450. doi: 10.1146/annurev-arplant-042110-103854

Moreira, X., Abdala-Roberts, L., Gols, R., and Francisco, M. (2018). Plant domestication decreases both constitutive and induced chemical defences by direct selection against defensive traits. *Sci Rep*, 8, 12678. doi: 10.1038/s41598-018-31041-0

Nakano, M., Shimada, T., Endo, T., Fujii, H., Nesumi, H., Kita, M., et al. (2012). Characterization of genomic sequence showing strong association with polyembryony among diverse Citrus species and cultivars, and its synteny with Vitis and Populus. *Plant Sci*, 183, 131–142. doi: 10.1016/j.plantsci.2011.08.002

Nathanson J. A. (1984). Caffeine and related methylxanthines: possible naturally occurring pesticides. *Science*, 226(4671), 184–187. doi: 10.1126/science.6207592

National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2022 June 14]. Available from: https://www.ncbi.nlm.nih.gov/

Nogata, Y., Sakamoto, K., Shiratsuchi, H., Ishii, T., Yano, M., and Ohta, H. (2006). Flavonoid composition of fruit tissues of citrus species. *Biosci Biotechnol Biochem*, 70(1), 178–192. doi: 10.1271/bbb.70.178

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Raab, T., López-Ráez, J. A., Klein, D., Caballero, J. L., Moyano, E., Schwab, W., et al. (2006). FaQR, required for the biosynthesis of the strawberry flavor compound 4-hydroxy-2,5-dimethyl-3(2H)-furanone, encodes an enone oxidoreductase. *Plant Cell*, 18(4), 1023–1037. doi: 10.1105/tpc.105.039784

Rao, M. J., Zuo, H., and Xu, Q. (2020). Genomic insights into citrus domestication and its important agronomic traits. *Plant Commun.*, 2(1), 100138. doi: 10.1016/j.xplc.2020.100138

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. doi: 10.3389/fpls.2019.01167

Schmelz, E. A., Huffaker, A., Sims, J. W., Christensen, S. A., Lu, X., Okada, K., et al. (2014). Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J*, 79(4), 659–678. doi: 10.1111/tpj.12436

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. doi: 10.3389/fpls.2015.00135

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. doi: 10.1016/j.plantsci.2019.110288

Strazzer, P., Spelt, C. E., Li, S., Bliek, M., Federici, C. T., Roose, M. L., et al. (2019). Hyperacidification of Citrus fruits by a vacuolar proton-pumping P-ATPase complex. *Nat. Commun.*, 10(1), 744. doi: 10.1038/s41467-019-08516-3

Su, C. F., Wang, Y. C., Hsieh, T. H., Lu, C. A., Tseng, T. H., and Yu, S. M. (2010). A novel MYBS3-dependent pathway confers cold tolerance in rice. *Plant Physiol*, 153(1), 145–158. doi: 10.1104/pp.110.153015

Swingle W.T., and Reece P.C. (1967). "History, world distribution, botany, and varieties" in The Citrus Industry, revised 2nd, vol 1, ed. Reuther W., Webber H.J. and Batchelor L.D. (Berkeley: University of California Press), 190–430.

Tadeo, F., Terol, J., Rodrigo, M.J., Licciardello, C., and Sadka, A. (2020). " Fruit growth and development" in The Genus Citrus, ed. Talon, M., Caruso, M. and Gmitter F.G., (Woodhead Publishing), 245-269.

Talon, M., Wu, G.A., Gmitter, F.G., Jr., and Rokhsar, D. (2020). "The origin of citrus" in The Genus Citrus, ed. Talon, M., Caruso, M. and Gmitter F.G., (Woodhead Publishing), 9–31.

Tanaka, T. (1954) Species problem in Citrus. Japanese Society for Promotion of Science.

Terol, J., Nueda, M. J., Ventimilla, D., Tadeo, F., and Talon, M. (2019). Transcriptomic analysis of Citrus clementina mandarin fruits maturation reveals a MADS-box transcription factor that might be involved in the regulation of earliness. BMC *Plant Biol*, 19(1), 47. doi: 10.1186/s12870-019-1651-z

Tholl, D., Sohrabi, R., Huh, J. H., and Lee, S. (2011). The biochemistry of homoterpenes--common constituents of floral and herbivore-induced plant volatile bouquets. *Phytochemistry*, 72(13), 1635–1646. doi: 10.1016/j.phytochem.2011.01.019

UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res., 49(D1), D480–D489. doi: 10.1093/nar/gkaa1100

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls:

the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, 43(1110), 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43

Wang, J., and De Luca, V. (2005). The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. *Plant* J, 44(4), 606–619. doi: 10.1111/j.1365-313X.2005.02552.x

Wang, L., He, F., Huang, Y., He, J., Yang, S., Zeng, J., Deng, C., Jiang, X., Fang, Y., and Wen, S. (2018). Genome of wild Mandarin and domestication history of Mandarin. *Mol. Plant*, 11:1024–1037. doi: 10.1016/j.molp.2018.06.001

Wang, L., Huang, Y., Liu, Z., He, J., Jiang, X., He, F., et al. (2021). Somatic variations led to the selection of acidic and acidless orange cultivars. *Nat Plants*, 7(7), 954–965. doi: 10.1038/s41477-021-00941-x

Wang, X., Xu Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., et al. (2017). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat Genet*. May;49(5):765-72. doi: 10.1038/ng.3839

Wei, Q. J., Ma, Q. L., Zhou, G. F., Liu, X., Ma, Z. Z., and Gu, Q. Q. (2021). Identification of genes associated with soluble sugar and organic acid accumulation in 'Huapi' kumquat (Fortunella crassifolia Swingle) via transcriptome analysis. *J Sci Food Agric*, 101(10), 4321–4331. doi: 10.1002/jsfa.11072

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686, doi: 10.21105/joss.01686

Wu, A., Allu, A. D., Garapati, P., Siddiqui, H., Dortay, H., Zanor, M. I., et al. (2012). JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in Arabidopsis. *Plant Cell*, 24(2), 482–506. doi: 10.1105/tpc.111.090894

Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., et al. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotechnol*. 32 (7):656-62. doi: 10.1038/nbt.2906

Wu, G.A., Terol J., Ibanez, V., Lopez-Garcia, A., Perez-Roman, E., Borredá, C., et al. (2018) Genomics of the origin, evolution and domestication of citrus. 2018; *Nature*. 544:311–316. doi:10.1038/nature25447

Wu, G. A., Sugimoto, C., Kinjo, H., Azama, C., Mitsube, F., Talon, M., et al. (2021). Diversification of mandarin citrus by hybrid speciation and apomixis. *Nat Commun*, 12(1), 1-10. doi: 10.1038/s41467-021-24653-0

Yactayo-Chang, J.P., Tang, H.V., Mendoza, J., Christensen, S.A., and Block, A.K. (2020). Plant Defense Chemicals against Insect Pests. *Agronomy*, 10, 1156. doi: 10.3390/agronomy10081156

Yang, X., Li, H., Yu, H., Chai, L., Xu, Q., and Deng, X. (2017). Molecular phylogeography and population evolution analysis of Citrus ichangensis (Rutaceae). *Tree Genetics & Genomes*, 13(1), 1-16. doi: 10.1007/s11295-017-1113-4

Supplementary Data

Supplementary Material

Gene Expression in Inedible Wild Ichang Papeda versus Acidic Sun Chu Sha Kat Mandarin

Lipids and Fatty Acids. The data (Supplementary Table 3) suggest that the expression levels of genes implicated in the regulation of fatty acid biosynthesis (Supplementary Figure 4.008) and elongation (Supplementary Figure 4.009) pathways, do not drastically differ between both species. However, the KEEG mapping revealed several interesting observations (Supplementary Table 2). One of the most relevant changes, for instance, was related to the upregulation in the papeda of practically all genes controlling fatty acid degradation (Supplementary Figure 4.071), very likely providing higher amounts of Acetyl-CoA to fuel the citrate cycle. It is also noticeable that in this species, 10 out of 11 DEGs regulating steroid biosynthesis, including lysosomal acid lipase/cholesteryl ester hydrolase (LOC18051398, EC:3.1.1.13, Supplementary Figure 4.012), that renders cholesterol, were downregulated, consistently with the repression of both squalene monooxygenase (LOC18033947 and LOC18033838, EC:1.14.14.17, Supplementary Figure 4.012) and the synthesis of brassinosteroids (Supplementary Figure 6). Another interesting observation in the sphingolipid metabolism is that in the papeda, DEGs are also mostly repressed except the conversion of ceramide to ceramide-P, a reaction catalyzed by ceramide kinase (LOC18031894, EC:2.7.1.138, Supplementary Figure 4.054, Supplementary Table 2), that appears to be participating in cold tolerance. Similarly, the synthesis of glycerol and triacylglycerol do not appear to be favored in ICH, in contrast to that of glycerol 3P and diacyl-glycerol 3P (Supplementary Figure 4.047, Supplementary Figure 4.049). In the peroxisome, two genes, acyl-CoA oxidase (LOC18036670 and LOC18047109, EC:1.3.3.6, Supplementary Figure 4.115) and a long-chain acyl-CoA synthetase (LOC18042917 and LOC1805469, EC:6.2.1.3, Supplementary Figure 4.115), acting in the ß-oxidation of fatty acids were upregulated. In other aspects of the lipid metabolism related to linoleic acid (Supplementary Figure 4.052), glycosphingolipids (Supplementary Figure 4.055 and Supplementary Figure 4.056), and ether lipid (Supplementary Figure 4.050), relevant observations were not obvious. Other results focused in glycerophospholipids and inositol phosphates, wax,

and jasmonate are described in the previous parts devoted to membranes and transduction, cell wall, and hormones respectively.

Gametophyte. In this category, more than two thirds of 46 genes, mainly transcription factors and receptors implicated in processes related to fertilization, karyogamy or the formation of gametophytes, were expressed at lower levels in ICH (Supplementary Table 3).

Organ Development, Differentiation. This group comprised 60 genes (Supplementary Table 3), in general transcription factors, that are well known to play relevant roles in the regulation of growth and development of practically all plant organs, such as leaf, vascular tissue, seed, stoma, flower, meristem or root. Some of these genes control directional growth, cell fate, meristem elongation, or aerial development. As many as 50 genes of this set were downregulated in ICH, including 11 genes related to meristem growth and DA1 (LOC18040667) a major regulator of final seed size.

Chloroplast. This category was clearly downregulated (94/20) in ICH, since many genes involved in specific processes listed in Supplementary Table 3, such as photosynthesis, the Calvin cycle, and the synthesis of chlorophyll or starch, showed lower number of transcripts in this species. In addition, the mapping of genes to the KEGG pathways revealed that all DEGs involved in Photosystems II and I, the photosynthetic electron transport, including ferredoxin--NADP+ reductase, petH (LOC18037401, EC:1.18.1.2) and several subunits of the F-type H+-transporting ATPase, ATPF1G (LOC18050746 and LOC18055418), and all components of the light harvesting chlorophyll protein complex), were unequivocally downregulated (Supplementary Figure 4.015 and 4.016; Supplementary Table 2). The synthesis of both chlorophylls, a and b, was also repressed (Supplementary Figure 4.071, Supplementary Table 2).

Cell division. The comparison between both species also revealed that categories containing genes directly related to cell division were essentially downregulated. The category (Supplementary Table 3) included 93 genes, 76 of which were downregulated. It was subdivided in different groups, e.g. Cell Cycle, that mostly contained cyclins (16/7), Cytokinesis, genes involved in the orientation and formation of the phragmoplast and cell plate (10/2), Mitosis, genes primarily related to the mitotic spindle formation, and other involved in the regulation of the G1 phase of the cell cycle, microtubule-

kinetochore conjugation, chromosome condensin complex or the control of mitotic checkpoints (22/3), DNA repair, comprising mostly transcripts coding for components of the double-strand break repair complexes, like the post-replicative DNA mismatch repair system (MMR) or the SMC5-SMC6 complex (7/3) and DNA replication, genes implicated in several complexes controlling DNA replication, such as the origin recognition complex (ORC), the GINS complex, the protein A complex (RPA/RP-A) or the cohesin complex (17/0). In addition, virtually all DEGs that mapped to KEGG pathways with roles on DNA replication (Supplementary Figure 4.099), base and nucleotide excision repair (Supplementary Figure 4.102, Supplementary Figure 4.103), mismatch repair (Supplementary Figure 4.104) and homologous recombination (Supplementary Figure 4.105) were downregulated.

Meiosis. This category (Supplementary Table 3) included mostly downregulated genes (9/1), that are required for normal meiosis, including meiotic recombination and crossovers, pairing of homologous chromosomes, formation of tetrad of microspores or chromosome segregation.

Cytoskeleton. Another category related to growth that exhibited higher number of down- (34) than upregulated genes (11), was Cytoskeleton (ST 2832), that included genes associated with the regulation of Actin (3/5) and Microtubule organization (26/8) and hence, with the cytoskeletal organization and cell morphology. The majority of the transcripts included in the microtubule group corresponded to motor proteins of the kinesin-like protein KIN type (24/2), and practically all of them, were downregulated (Supplementary Table 3).

Cell Wall. The category included 74 genes that were classified in three groups, genes mostly coding for enzymes implicated in the own synthesis of the components and structural proteins of the cell wall (57), genes related to the extracellular Matrix (7), and genes implicated in the formation of the Cuticular wax (10). In the first group of genes, downregulation was in general more prominent that upregulation (34/23) a circumstance that was observed for cellulose synthases, extensins, arabinogalactan proteins, arabinosyltransferases, expansins, polygalacturonases, pectinesterases, xyloglucan acetyltransferases, xyloglucan glycosyltransferase and other transcripts implicated in secondary wall formation. In contrast, pectinesterase inhibitors, ligases, glycosyltransferases, galacturonosyltransferases and xyloglucan endotransglucosylases (XET) were in general up. In the second and third groups, involving genes related to the

extracellular matrix (cell surface adhesion proteins and signaling receptors) and to the cuticular wax (ABC transporters of the G family, lipid-transfer proteins, and alkane hydroxylases), respectively, no clear differences between the abundance of down and up regulated genes were observed (Supplementary Table 3). In addition, the papeda showed downregulation (Supplementary Figure 4.011, Supplementary Table 2) of a fatty acid omega-hydroxy dehydrogenase, ACE, (LOC18036702, 1.1.-.-), involved in the formation of oxo fatty acids in the suberin biosynthesis, and of the aldehyde decarbonylase CER1 gene (LOC18049734, EC:4.1.99.5, associated with the production of wax.

Cell Expansion. The number of down and up regulated genes in this category was only slightly different (8/11). The category mainly contained regulatory genes of cell expansion, e.g., growth-regulating factors, receptor-like serine/threonine-protein kinases, or wall-associated receptor kinases. For example, one of the downregulated genes, was protein LONGIFOLIA 1 (LOC18048398, Supplementary Table 3), that promotes longitudinal polar cell elongation.

Receptors and Protein Kinases. The 108 DEGs include in this category (Supplementary Table 3) were those not assigned to any specific cellular process, out of which 63% were expressed at lower levels in the papeda.

Membrane, ER and Signalling. Members of this category included genes related to the Rho GTPase signaling complex, calcium sensors and signaling transduction and other second messengers, such as cAMP, inositol triphosphate and diacylglycerol, phosphatidylinositol 3-kinase signaling, signal recognition particle, and genes involved in the synthesis of phosphatidylcholine and other membrane components (Supplementary Table 3). In this category downregulation was not so prominent (40/33), although some gene families like that of proteins IQ-DOMAIN 1 and 14 were manly expressed at lower levels (8/2). The synthesis of membrane glycerophospholipids, such as phosphatidylethanolamines and phosphatidylcholines, also appears to be repressed in the papeda (Supplementary Figure 4.049). In the inositol phosphate metabolism (Supplementary Figure 4.048, Supplementary Figure 4.108, Supplementary Table 2), ICH showed downregulation of genes coding for inositol mono-, bis-, tris-, pentakis-, and hexaphosphate (phytate). However, the 2 last steps in the synthesis of inositol-1,3,4P3 (IP3) which encompass 1-phosphatidylinositol-4-phosphate 5-kinase, PIP5K (LOC18042642, EC:2.7.1.68), that generates phosphatidyl

inositol-4,5 biphosphate, and phosphatidylinositol phospholipase C, gamma-1, PLCD (LOC18044207, LOC18043460 and LOC18044208, EC:3.1.4.11), that hydrolyzes this last compound to diacylglycerol and IP3, were upregulated, suggesting that the availability of these two second messenger molecules is superior in ICH.

Chromatin, Histones. This category was composed of 27 genes mostly related to histones, chromatin assembly and remodeling factors, and histone methyltransferases. In this group, 22 genes, including 13 transcripts coding for the 4 core components of nucleosome and the chromatin fiber, were downregulated (Supplementary Table 3).

Transcription. In this category, as many as 75 genes out of 118 were downregulated. Most DEGs included in this category were transcription factors (54/36) and the rest of members were divided in 2 groups, Post-transcription including splicing (5/1) and gene silencing (6/2), and RNA polymerase (10/4) (Supplementary Table 3). The KEGG mapping showed that in general, minor differences were observed among the basal transcription factors for RNA polymerase II (Supplementary Figure 4.098) and the core and specific subunits of the polymerases (Supplementary Figure 4.097). In addition, most DEGs included in nucleocytoplasmic transport (9/3; Supplementary Figure 4.094) and mRNA surveillance pathway (9/4, Supplementary Figure 4.095) were downregulated and those implicated in RNA degradation were upregulated (Supplementary Figure 4.096).

Nucleic Acids Processing. The category was populated to a large degree with helicases, DNA methyl transferases, ligases, endonucleases, ribonucleases, primases and other factors acting on nucleic acids. Downregulated genes were predominant in the papeda, since 19 genes out of 27 were expressed at lower levels (Supplementary Table 3).

Translation. The category was dominated by upregulated genes (18/25) that mostly corresponded to the group of 40 S and 60 S Ribosomal proteins, and genes implicated in ribosome biogenesis and other factors controlling translation initiation and the termination of nascent peptide synthesis (Supplementary Table 3).

Trafficking, Vesicles. The category was characterized by a higher number of upregulated genes (9/16), mostly involved in the secretion of vesicles from ER, via the Golgi up to the trans Golgi network, regulation of clathrin and non-snare proteins, autophagy, endosome trafficking and recycling, and in the processes of micropinocytosis, endocytosis and exocytosis (Supplementary Table 3). A set of

upregulated genes representative of this category was that of the GTP-binding proteins generally associated with the regulation of membrane trafficking, for instance, SAR1 (LOC112096942, Supplementary Table 3 and LOC18050081, EC:3.6.5.-, Supplementary Figure 4.112, Supplementary Table 2), a GTPase regulating the formation of Golgy bodies. Other upregulated DEGs involved in protein processing in the ER represented in Supplementary Figure 4.112 (Supplementary Table 2) were Sec61 (LOC18052999) and TRAP (LOC18035022), components of the translocon, a complex that translocates polypeptides, or several genes of the ERAD (ER-associated degradation), such as Hsp40 (DNAJA2, LOC18052272), NEF (HSPBP1, LOC18039711), Ufd1 (LOC18033916) and DUB (ATXN3, LOC18038522), a previous step to the proteasome. Upregulated genes involved in ER export included signal peptidases, translocation channels and related proteins of the Sec dependent pathway (Supplementary Figure 4.101, Supplementary Table 2). Furthermore, upregulated genes implicated in endocytosis were represented by clathrins, Rab proteins (small G proteins attached to vesicles), VPS (vacuolar protein sorting) and CHMP4 (charged multivesicular body proteins) (Supplementary Figure 4.113, Supplementary Table 2).

Ubiquitination. The category was comprised mostly by upregulated transcripts (26/38; Supplementary Table 3) involved in Proteasome, specifically major ubiquitin players such as the E3 ubiquitin-protein ligase complexes (13 genes, CUL3-RBX1-BTB; Elongin BC-CUL2/5-SOCS-box protein), or the E2s enzyme SCF (SKP1-CUL1-F-box protein)-type complexes (3 genes), in addition to chaperones and chaperonins, polyubiquitins and proteasome subunits. Supplementary Figure 4.112 shows additional upregulated genes located in the ER membrane or in the cytoplasm, belonging to the ubiquitin ligase complexes, while Supplementary Figure 4.110 illustrates that practically all 9 reactions describing the E2 (ubiquitin-conjugating enzyme) were upregulated.

Plant Defense. This category containing the highest number of DEGs (263), was enriched with downregulated genes (150) displaying a wide range of disease resistance proteins against pathogens (Supplementary Table 3). Among them, some are classified as general defense proteins involved in immunity, hypersensitive response, or pathogen responses, while others are defined by anti-fungal, -bacterial, and -viral activities.

Cofactors and Vitamins. DEGs implicated in the synthesis of cofactors and vitamins are listed in Supplementary Table 3, and other genes related to this category involved in

different pathways are presented in Supplementary Figure 4.062 and Supplementary Figure 4.064-4.070. In general, the number of DEGs in these processes is too short to extract valuable remarks, although maybe is worth to mention that the synthesis of pantothenate (vitamin B5) and CoA, riboflavin and flavin mononucleotide, nicotinamide adenine dinucleotide, and the compounds integrating vitamin B6 show a tendency towards upregulation in the papeda.

DEGs included in the categories **Cell Death** (8/2, genes regulating programmed cell death, for instance, aspartic proteinases, co-chaperones…), **Light Signaling** (13/13, mostly transcription factors and photoreceptors involved in light-signal transduction), **Seed, Embryo Development** (5/10, seed storage, LEA proteins and embryo development) and **Flowering** (13/12, transcription factors, receptors and proteins related to FT controlling the flower-promoting signal), in principle do not appear to provide unambiguous information useful to accomplish the objectives proposed in this work (Supplementary Table 3).

The KEEG mapping also suggested that the amino sugar/nucleotide sugar metabolism (Supplementary Table 2) was also apparently more active in ICH than in SCM (Supplementary Figure 4.045). For instance, last regulatory steps controlling the synthesis of major intermediates participating in carbon metabolism, such as mannose-6P, UDP-L-arabinose, UDP-glucuronic acid, UDP-glucose, UDP-rhamnose, UDP-galactose and pectin were unequivocally upregulated. However, there were not too many differences in the expression levels of genes participating in N-glycan biosynthesis (Supplementary Figure 4.043) or glycosaminglycan degradation (Supplementary Figure 4.046).

Gene Expression in Palatable Pummelo/Mandarin Genetic Admixtures versus Acidic Sun Chu Sha Kat Wild Mandarin

Raw expression data in this section are presented in Supplementary Table 5, although the number of genes included in the below categories are in general too short to support a solid enrichment analysis.

Plant defense. The most populated category was Plant Defense (63), that surpassed even the category of uncharacterized process including genes with undetermined physiological roles (39). In comparison with SCM, the number of genes down- and upregulated in the three segregant mandarins was identical (31/32). The majority of

these genes are classified as disease resistance proteins, that guard the plant against pathogens containing an avirulence protein, such as DSC1 like proteins (5/4), TMV resistance proteins N (3/0), or the RGA gene family (4/9). Several of them recognize fungal effectors and elicitors of P. infestans, P. syringae, T. viride or C. fulvum, modulate herbivory resistance, or act as triggers of hypersensitive response (HR). The three TMV resistance protein N detected, for instance, were downregulated in the palatable mandarins, as several genes with general responses and central roles in resistance, such as RPM1-interacting protein 4 (LOC18033466), ABC transporter G family member 34 or HAK1 (LOC18039121). Conversely, all 4-disease resistance protein At4g27220 (LOC18040211, LOC112097691, LOC18042654 and LOC18041183) and osmotin-like protein (LOC18035225) were upregulated in segregants. In this category other upregulated genes are glucan endo-1,3-β-glucosidases (LOC112101121 and LOC1803342) with roles in plant defense against fungal pathogens and β-glucosidase 17 (LOC18043267), regulating the synthesis of scopoletin/scopoline, coumarins that have deterrent effect on herbivores.

Abiotic stress. Most DEGs in the Abiotic stress category (6/1) were downregulated in the three segregants.

Amino acid metabolism. Most DEGs in the Amino acid metabolism category (5/1) were downregulated in the three segregants. Aminotransferases involved in amino acid degradation were predominant.

Cell Division. This category encompassed 6 genes which showed a down to up rate of 0.5.

Cell Wall. This group contained relevant genes (8/8) controlling major structural components of cell wall. Downregulated genes were cellulose synthase-like protein E1 (LOC18055183), expansin-like B1 (LOC18048704, Supplementary Table 5), pectin acetylesterase 8 (LOC18040742), pectinesterase/pectinesterase inhibitors (LOC18053206, LOC112095422 and LOC18040524), protein trichome birefringence-like 8 (LOC18038166, Supplementary Table 5), or cytochrome P450 84A1, which is a ferulate 5-hydroxylase involved in phenylpropanoids metabolism (LOC18036838), involved in lignin biosynthesis, were all downregulated.

Transcription. This category was populated with transcription factors and RNA polymerases and related factors, most of them repressed (10/2).

Receptors, Kinases, Transduction. The category was composed of 14-3-3 proteins, F-box/FBD/LRR-repeat proteins, and serine/threonine-protein kinases. Upregulated genes (4/12) were predominant.

Translation. The number of up and downregulated transcripts related to tRNA and ribosome biogenesis in this category were similar (5/3),

Ubiquitination. This category containing a similar number of up and down regulated genes (4/3), implicated in proteasomal degradation.

Membranes, ER, and Signaling. The group of genes (6/4) included in this category, such as adenylate cyclases, potential calcium sensors, or components of the GTPase signaling complex, did not showed a prevalent expression pattern.

Redox. DEGs clustered in the Redox category (5/5), mostly cytochromes, L-ascorbate oxidase, and probable glutathione S-transferases, contained a similar number of up and down regulated genes. In this category, the senescent protein SRG1 (LOC18054377) regulating plant immunity was repressed in the three palatable mandarins.

Trafficking, Vesicles. Synthesis of N-glycans was probably downregulated since α-mannosidase (LOC18036311), the first committed step of this pathway was repressed, although several other genes included in this category (3/5) were mainly upregulated.

Organ development, Differentiation. This category included several 3-oxo-Delta(4,5)-steroid 5-β-reductases involved in xylem and phloem pattern formation, that were mostly upregulated (1/3, LOC18041865, LOC18053615, LOC18041176 and LOC18041175).

Chloroplast. Flowering. (4/0), and Gametophyte (5/1) These were tree categories focused on growth and development, showing mostly downregulation, (13/1, 4/0, and 5/1, respectively).

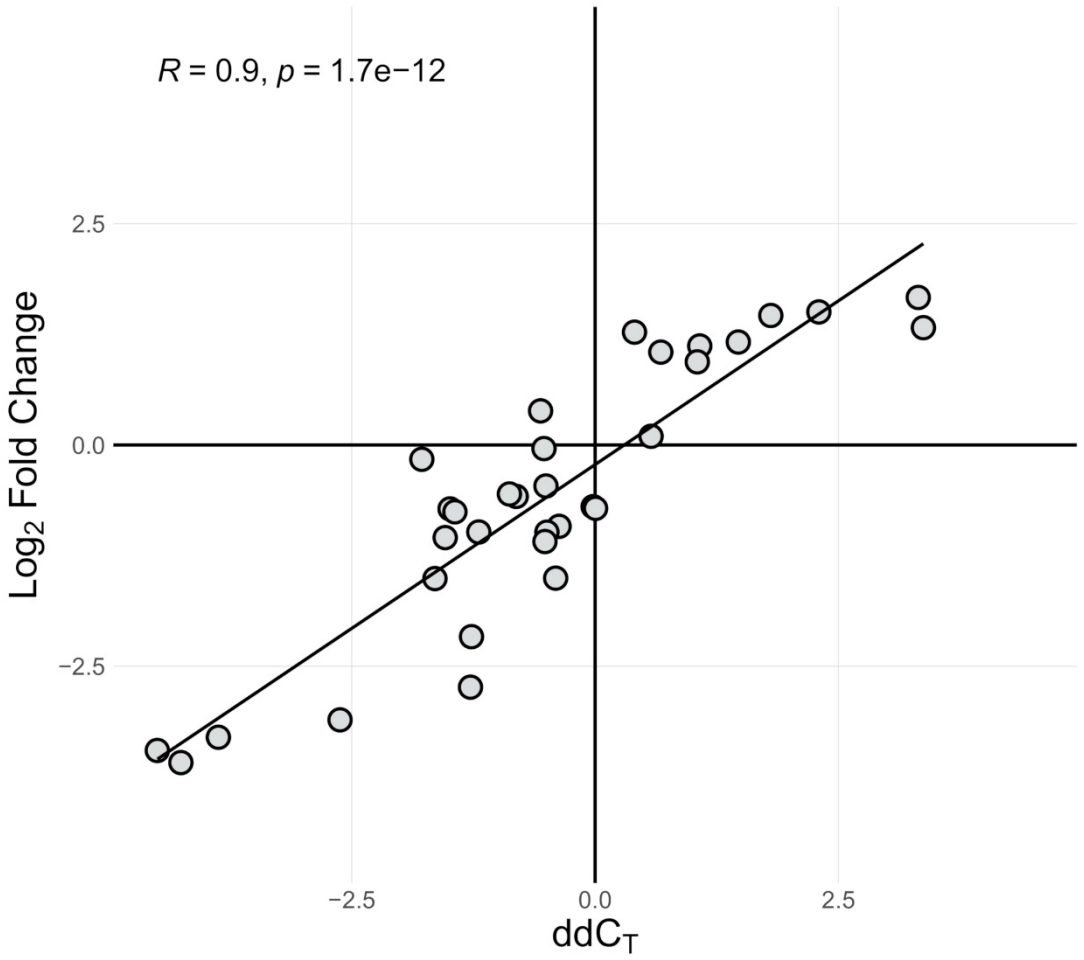
Hormonal Regulation. In this category (14/7), most DEGs related to ABA, auxins, cytokinins, ethylene, jasmonic acid, polyamines and salicylic acid were downregulated.

Cell Death. Light Signaling. Meiosis. Phytoalexins. S-adenosylmethionine Metabolism. Nucleobases. Seed, Embryo, Development. For these categories, the number (1/2, 3/2, 0/1, 0/1,0/2, 1/3 and 2/2) or relevance of the DEGs was too scarce to suggest unambiguous comments.

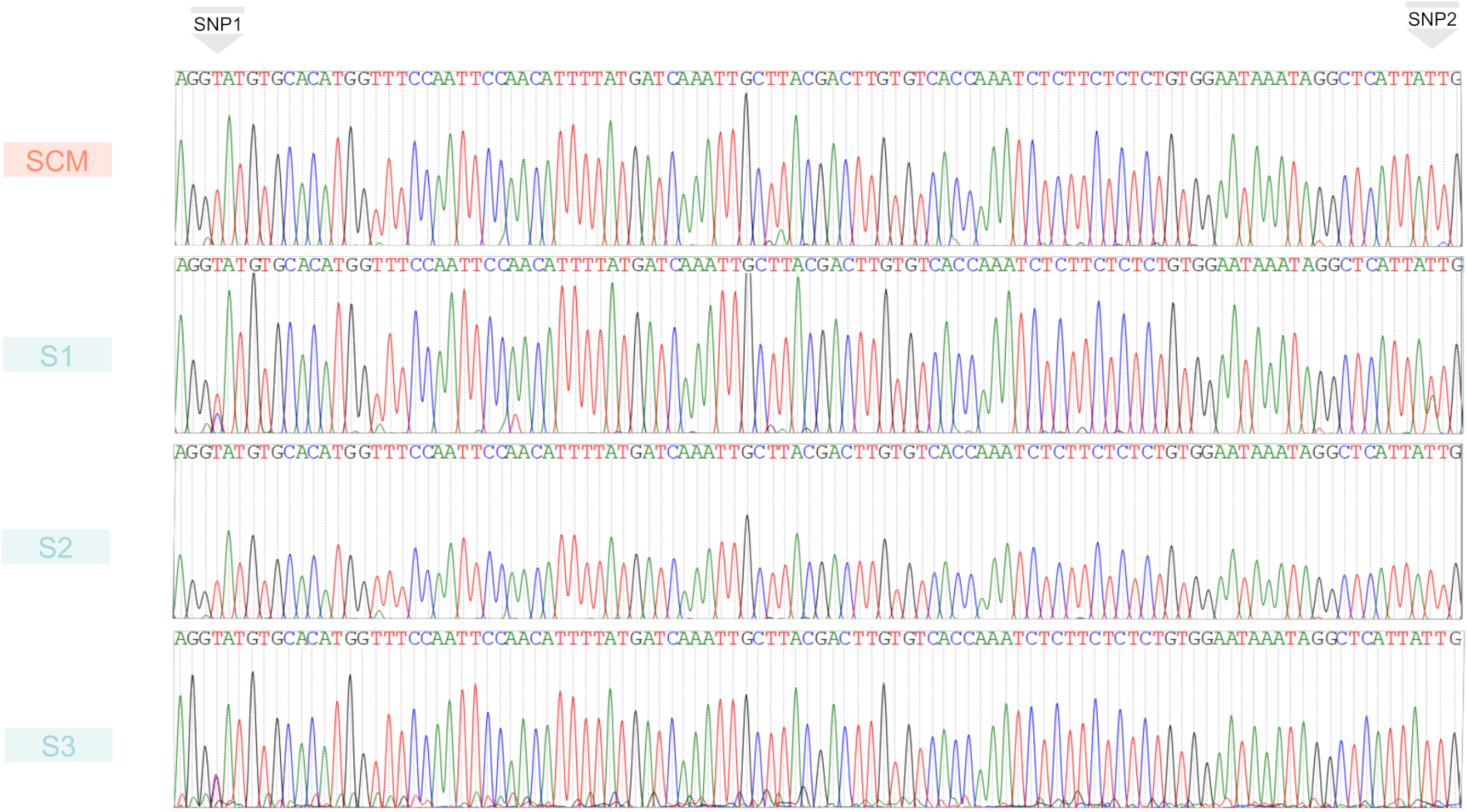
Lipids and fatty acids. Among the upregulated genes in this category (5/7), there were two genes implicated in elongation of fatty acids (LOC18048831 and LOC18033122), a 3-ketoacyl-CoA synthase 1 (LOC18048831) relevant for the synthesis of cuticular wax and suberin, and an elongation of fatty acids protein A-like (LOC112098405). Other two genes coding for peroxisomal enzymes (LOC18050645 and LOC18055962) participating in the long-chain fatty acid catabolic pathways were also upregulated. Repressed genes were an additional 3-ketoacyl-CoA synthase 5 (LOC18033226), CDP-diacylglycerol--inositol 3-phosphatidyltransferase 1-like (LOC112096415), involved in phospholipid biosynthetic process and sterol 3-β-glucosyltransferase UGT80A2 (LOC18039007) controlling the biosynthesis of sterol glucosides, the most abundant sterol derivatives in higher plants, especially in seeds.

Flavonoids. This category contained similar number od down and up regulated genes (6/6). Although three flavonoid 3'-monooxygenases (LOC18032922, LOC18052870 and LOC18053379) were indistinctly down and up regulated in the three segregants, two central regulatory genes of the pathway, chalcone-flavonone isomerase (LOC18043493) and naringenin,2-oxoglutarate 3-dioxygenase (LOC18036490) were upregulated. Anthocyanidin 3-O-glucosyltransferase (LOC18039058) and anthocyanidin 5,3-O-glucosyltransferase (LOC18032110), participating in the biosynthesis of anthocyanins were also similarly upregulated. Further metabolism appears to be downregulated, for instance, the reduction of S-glutathionylquercetin to quercetin (LOC18043062), the conversion of phloretin to phlorizin (LOC18039246), the biosynthesis of lignans (LOC18041854), phenylpropanoid monomers of the lignin polymers, or the acylation of the anthocyanin glucose (LOC18032737). As mentioned in the Results section of the manuscript, mandarins do not contain anthocyanins.

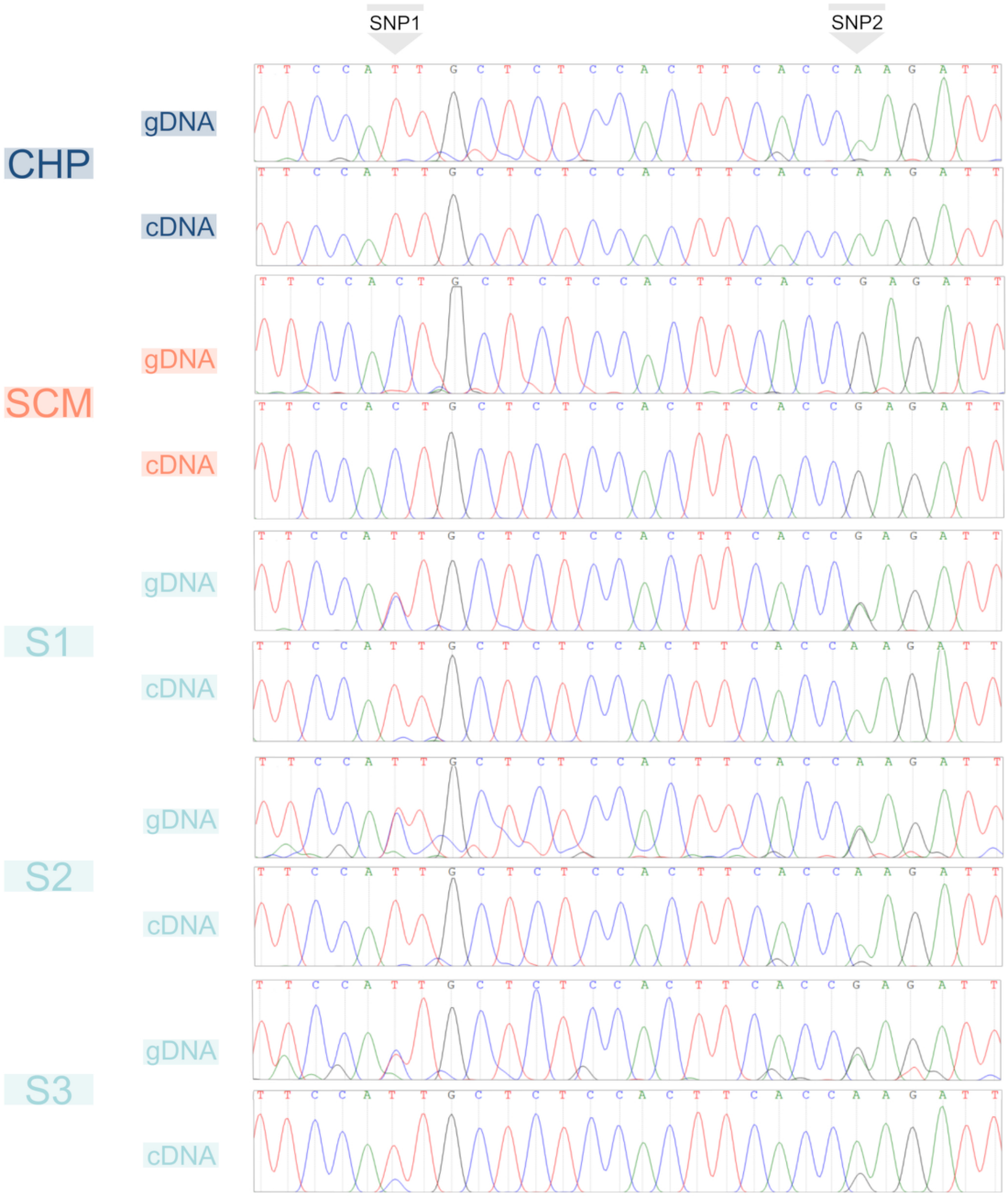
Supplementary Figures



Supplementary Figure 1. Linear correlation between RNA-seq Log_2 FC and RT-qPCR relative expression (ddCT) of gene/sample combinations tested in the study.



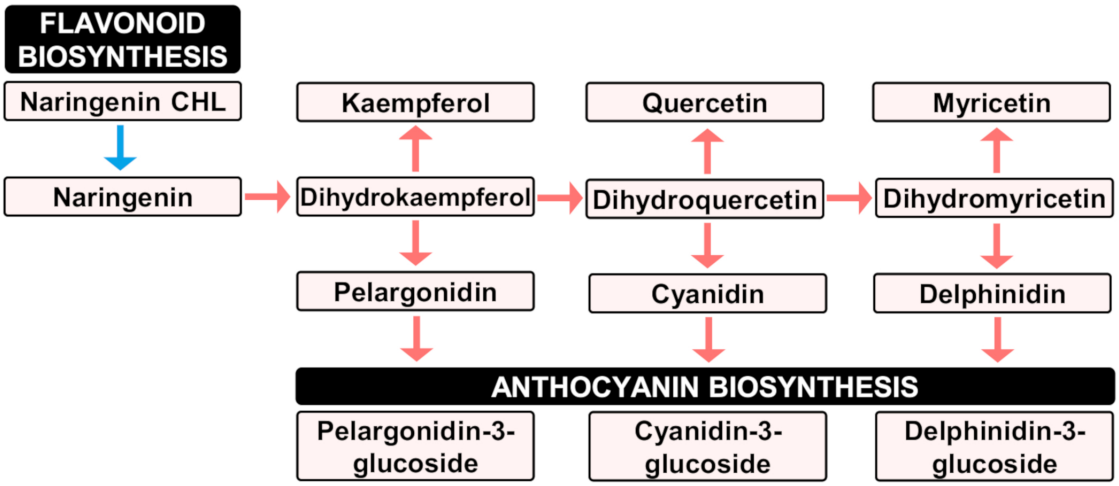
Supplementary Figure 2. Validation of an admixture pattern shift, a crossover in CHR2. Chromatograms correspond to PCR amplicons of the target region. S1 = PU/MA for both SNP1 and SNP2. S2 = MA/MA for both SNP1 and SNP2. S3 = PU/MA for SNP1 and MA/MA for SNP2.



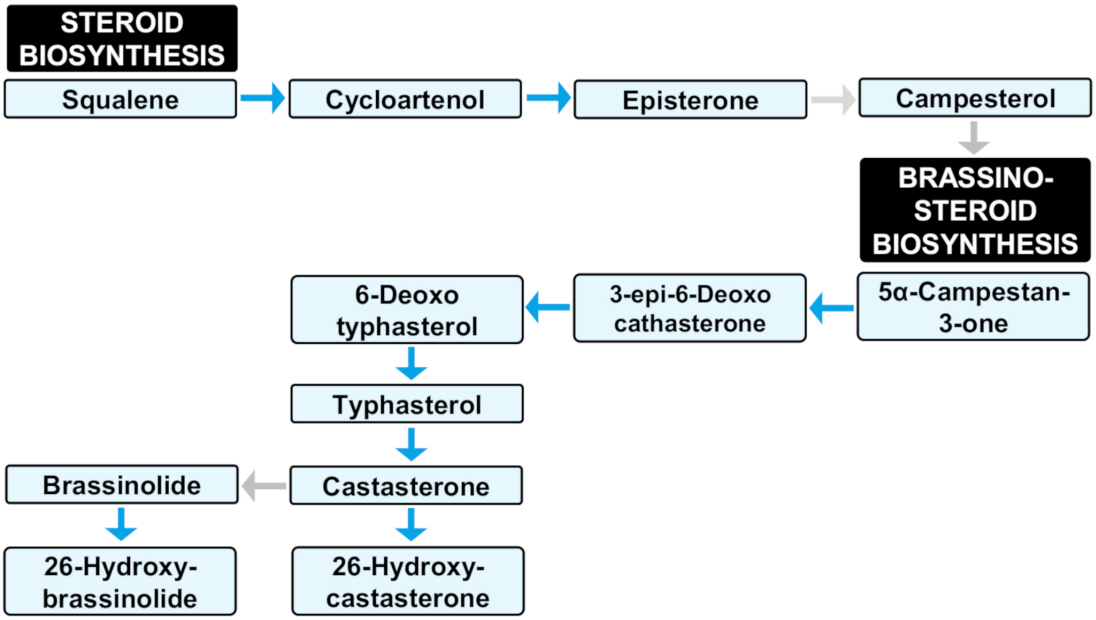
Supplementary Figure 3. Validation of the allele differential expression of LOC18040524. For SNP1, gDNA of CHP is T/T and gDNA of SCM is C/C. gDNA of the three segregants is T/C, while cDNA is T/T. For SNP2, gDNA of CHP is A/A and gDNA of SCM is G/G. gDNA of the three segregants is A/G, while cDNA is A/A.

See files in *Supplementary_Figure_4.zip*. They are available in the online version at https://doi.org/10.3389/fpls.2022.982683

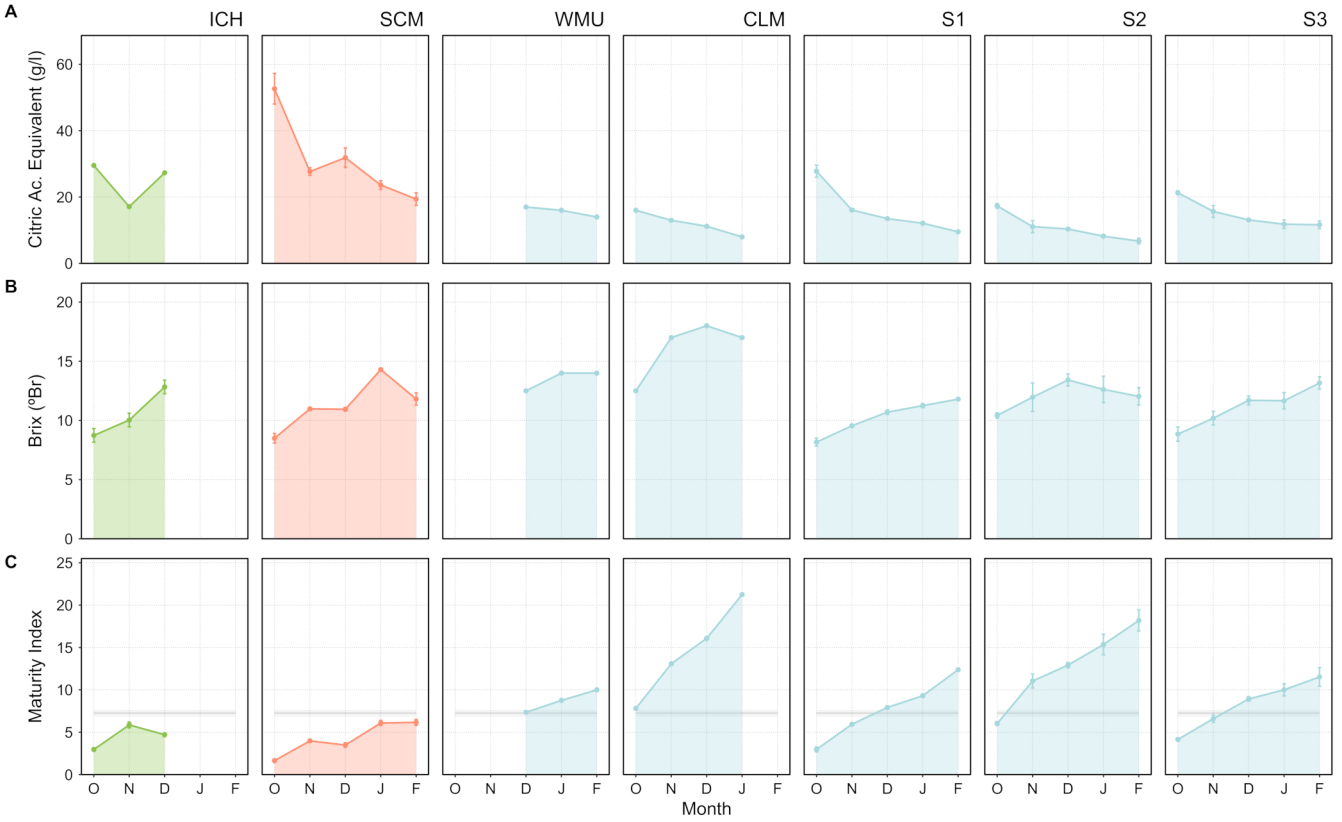
Supplementary Figure 4. Representation of 2342 DEGs in the pulp of developing fruitlets of ICH as related to SCM (Supplementary Table 3), mapping to 117 pathways as defined in the KEGG database resource (Kanehisa et al., 2016). Differential expression is represented by three uniform colors (up = red; blue = down; yellow = undetermined) without indication of the Log_2FC. "Undetermined" = two or more genes share the same KEGG identifier, but show opposite expression trends, as long as one of the biological replicates reaches at least 100 reads.



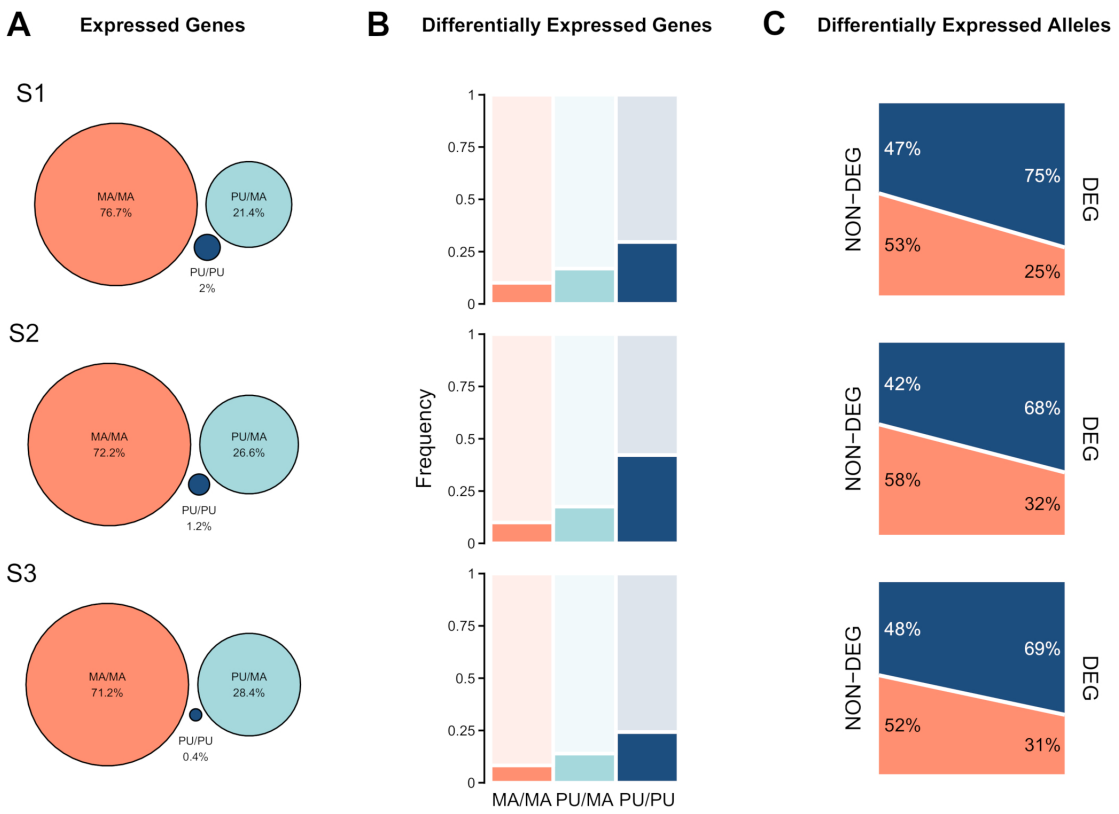
Supplementary Figure 5. Proposed activity of major flavonoid biosynthetic pathways, deduced from DEGs in the pulp of developing fruitlets of ICH as related SCM. Solid arrows indicate gene regulation: red = upregulation; blue = downregulation. Reaction products and pivotal pathways are embedded in colored and black boxes, respectively.



Supplementary Figure 6. Proposed activity of brassinosteroid pathways, deduced from DEGs in the pulp of developing fruitlets of ICH as related to SCM. Solid arrows indicate gene regulation: blue = downregulation, and grey = no differential expression. Reaction products and pivotal pathways are embedded in colored and black boxes, respectively.



Supplementary Figure 7. Pivotal biochemical parameters determining fruit quality, in ripening fruits of ICH (Ichang papeda), SCM (Sun Chu Sha Kat mandarin), WMU (W. Murcott mandarin), CLM (Clementine mandarin), and segregants S1, S2, and S3. **(A)** Acidity content measured as citric acid equivalents (g/l). **(B)**. Sugar content measured as total soluble solids (ºBr). **(C)** Maturity index computed as the ºBr/acidity ratio (g/100cm3). The gray line indicates the minimum value established by the European Commission Regulation on marketing standards in the fruit and vegetable sector. (https://www.boe.es/doue/2011/157/L00001-00163.pdf). Measurements for WMU and CLM fruits were retrieved from the variety record collection of the Instituto Valenciano de Investigaciones Agrarias (https://ivia.gva.es/va/variedades). Vertical bars represent standard error (n =3).



Supplementary Figure 8. Differential gene and allele expression in the pulp of developing fruitlets of segregants S1, S2 and S3, as related to SCM. Haplotypes: MA/MA = mandarin/mandarin (orange), PU/MA = pumelo/mandarin (light blue) and PU/PU = pumelo/pumelo (dark blue). **(A)** Percentage of expressed genes in each haplotype sequence. **(B)** Frequency of DEGs in each haplotype sequence. **(C)** Percentage of genes expressing either mandarin or pummelo alleles in non-DEGs and DEGs.

Supplementary Tables

Supplementary Table 1. List of primers used in the study.

Gene Name Abbrev.	Sample	*Forward*	*Reverse*
CitPH1	ICH;SCM;WMU;CLM;S1;S2;S3	CCAGAGACAGATGTGTCGTCA	CATGATTGCACGGTCCATGGT
CitPH5.2	SCM;WMU;CLM;S1;S2;S3	GGAGCTCTGTTGATGTGTGCA	ATGAACTTAATGACGTCCAATGGT
CitPH5.2	ICH	GGAGCTCTGTTGATGTGTGCA	ATGAACTTAATGACGTCCAATGGT
CitMAC9F1	ICH;WMU;S1;S2;S3	GCAGAAACGGATTGATGAAGAG	CATCAGGGACAACTTCCTCAG
CitMAC9F1	SCM	GCAGAAACGGATTGATGAAGAG	CATCAGGGACAACTTYCTCRG
CitAN1	SCM;WMU;CLM;S1;S2;S3	ACCGGTTATGATAGGTAGCAGT	GCTGCTTCTGAAGAATGGTAGA
CitAN1	ICH	CCCGGTTATGATAGGTAGCAGT	GCTGCTTCTGAAGAATGGTAGA
CitPH3	ICH;SCM;WMU;CLM;S1;S2;S3	CTGAGATTCTTAGTGATGGCTTC	CAGATGTCACTGGAGTTGTGC
CitPH4	SCM;WMU;CLM;S1;S2;S3	GATCTCATTCTTCGCCTACATC	CTTGGCTTATCAGCTTCTTACTC
CitPH4	ICH	GATCTCATTCTTCGCCTACATC	CTTGGCTTACCAGCTTCTTACTC
CitAN11	ICH;SCM;WMU;CLM;S1;S2;S3	GGTCTCCTGCACAGCCTGA	AGTTAGCTATCCCAAACGAGCA

CitPH5.1	SCM;WMU;CLM;S1;S2;S3	TCCTTGGTGACAGTGAGGGA	AAACTGCAGCTCTGATGGCT
CitPH5.1	ICH	TCCTTGGTGACAGTGAGGGA	AAGCTGCAGCTCAGATGGCT
CitSO	ICH;SCM;WMU;CLM;S1;S2;S3	CTTCTGGGAGAAAGCTTTTGAATG	CCAGAACAACATTGATAGGAATGA
CitERF13	SCM;WMU;CLM;S1;S2;S3	TCAGGCTCAGGCTAGTGGAT	GCGTATCATAGGTCCCCAGC
CitERF13	ICH	TCAGGCTCAGGCTAGTGGAT	GCGTATCGTAGGTCCCCAGC
CitVHA_C4	ICH;SCM;WMU;CLM;S1;S2;S3	CGACGAAACTGCTCCCTTCT	GATTGGCCTGCTCGAGATGA
ase.1.1	SCM;WMU;CLM;S1;S2;S3	CGGGCCACTTCACTCTCATT	AACACAGCTGGTCCTGCAAA
ase.1.2	SCM;WMU;CLM;S1;S2;S3	AGTCCAGCCATCAATGACACT	ACCTTGCCTGATTGCCACAA
gs.4.1	ICH;SCM;WMU;CLM;S1;S2;S3	GGAGGTGAAACTGGAAGAGGG	TGAGGGTGAAAGTCGGTTGA
gs.4.2	ICH;SCM;WMU;CLM;S1;S2;S3	GGCGGCGTACCAGATCATAA	CGCTCCAGTGACAATGTTGG
co.1	SCM;WMU;CLM;S1;S2;S3	GTCACCAATTCTGTCTCTGTGA	GCGGGAGGGCCAATATAGTC
CitACTIN11	CLM;S1;S2	GGYATTGCCGATAGAATGAGCA	TCATACTCAGCCTTTGCAATCCA
CitACTIN11	ICH;SCM;WMU;S3	GGCATTGCCGATAGAATGAGCA	TCATACTCAGCCTTTGCAATCCA
CitUBC	ICH;SCM;WMU;CLM;S1;S2;S3	GTGCAGCGAGAGAAATCAGC	ACTTGTGGAGGTTGCAGAGG

Supplemental Table 4. Expression of genes encoding last steps in the biosynthesis of amino acids and amino acid-derived compounds, in the pulp of developing fruitlets of ICH as related to SCM.

Product	Process	Metabolite			Expression
Amino acid	Biosynthesis	Glutamic acid	Glu	E	Up
Amino acid	Biosynthesis	Glutamine	Gln	Q	Up
Amino acid	Biosynthesis	Phenylalanine	Phe	F	Up
Amino acid	Biosynthesis	Proline	Pro	P	Up
Amino acid	Biosynthesis	Hydroxyprolin	Hyp	O	Up
Amino acid	Biosynthesis	Tyrosine	Tyr	Y	Up
Amino acid	Biosynthesis	Glycine	Gly	G	Up
Amino acid	Biosynthesis	β-Alanine	Ala		Up
Amino acid	Biosynthesis	Lysine	Lys	K	Non Diff
Amino acid	Biosynthesis	Histidine	His	H	Non Diff
Amino acid	Biosynthesis	Tryptophan	Trp	W	Non Diff
Amino acid	Biosynthesis	Cysteine	Cys	C	Non Diff
Amino acid	Biosynthesis	Methionine	Met	M	Non Diff
Amino acid	Biosynthesis	Aspartic acid	Asp	D	Non Diff
Amino acid	Biosynthesis	Serine	Ser	S	Non Diff
Amino acid	Biosynthesis	Arginine	Arg	R	Non Diff
Amino acid	Biosynthesis	Threonine	Thr	T	Non Diff
Amino acid	Biosynthesis	Asparagine	Asn	N	Non Diff
Amino acid	Biosynthesis	Valine	Val	V	Down
Amino acid	Biosynthesis	Isoleucine	Ile	I	Down
Amino acid	Biosynthesis	Leucine	Leu	L	Down
Amino acid	Degradation	Leucine	Leu	L	Up

Amino acid	Degradation	Valine	Val	V	Up
Amino acid	Degradation	Isoleucine	Ile	I	Up
Derived compound	Biosynthesis	Uracil	Ura	U	Up
Derived compound	Biosynthesis	Betaine			Up
Derived compound	Biosynthesis	Ornithine			Up
Derived compound	Biosynthesis	γ-aminobutyric acid	GABA		Up
Derived compound	Biosynthesis	Acetoacetic acid			Up
Derived compound	Biosynthesis	Indolacetic acid		IAA	Up
Derived compound	Biosynthesis	Pantothenic acid			Up
Derived compound	Biosynthesis	Taurine			Up
Derived compound	Biosynthesis	Hypotaurine			Up
Derived compound	Biosynthesis	Transcinnamic acid			Up
Derived compound	Biosynthesis	Homoserine			Non Diff
Derived compound	Biosynthesis	Spermidine			Down
Derived compound	Biosynthesis	Homocysteine			Down
Derived compound	Biosynthesis	Putrescine			Down
Derived compound	Biosynthesis	Dopamine			Down
Derived compound	Biosynthesis	Serotonine			Down
Derived compound	Biosynthesis	Tryptamine			Down
Derived compound	Biosynthesis	Selenocompounds			Down
Derived compound	Biosynthesis	Shikimate			Down
Derived compound	Biosynthesis	Shikimate-3P			Down
Derived compound	Biosynthesis	Ethylene			Down
Derived compound	Biosynthesis	Pyruvate			Down

Supplementary Table 6. Normalized read counts of relevant genes reported to be involved in acid regulation of citrus fruits (Aprile et al., 2011; Butelli et al., 2019; Strazzer et al., 2019; Huang et al., 2021), obtained in RNA-seq analyses of developing fruitlet pulp of ICH, SCM, and S1, S2, and S3 segregants.

			Normalized Read Count													
	Gene Id	**Gene Name**	**SCM.R1**	**SCM.R2**	**SCM.R3**	**S1.R1**	**S1.R2**	**S1.R3**	**S2.R1**	**S2.R2**	**S2.R3**	**S3.R1**	**S3R2**	**S3.R3**	**ICH.R1**	**ICH.R2**	**ICH.R3**
CitPH1	LOC18037376	magnesium-transporting ATPase, P-type 1	2329	2391	2232	119	417	231	1334	1108	702	1239	1246	1279	1590	1846	2655
CitPH5.2	LOC18035739	ATPase 10, plasma membrane-type	34387	31704	30654	1352	4465	2598	21860	17028	12920	19662	18552	19818	26028	27433	42348
CitMAC9F1	LOC18037289	uncharacterized LOC18037289	15690	14693	12000	725	2119	1217	13343	11010	7871	6648	6897	7351	5139	5698	7581
CitAN1	LOC18047507	basic helix-loop-helix protein A	7793	6735	6744	323	827	557	4032	3057	2496	3422	3383	3582	2589	2772	4061
CitPH3	LOC18038669	WRKY transcription factor 44	4114	3812	3798	1231	1481	1384	2743	2338	1988	2268	2410	2301	2375	2506	3073
CitPH4	LOC18053295	transcription factor MYB34	2977	2698	2560	140	429	224	1392	1258	858	1080	1141	1207	1518	1767	2472
CitAN11	LOC18032473	protein TRANSPARENT TESTA GLABRA 1	583	558	502	511	460	490	871	693	756	673	613	736	647	728	903
CitPH5.1	LOC18035736	ATPase 10, plasma membrane-type	482	498	499	45	95	70	126	113	72	102	191	190	1076	1127	1674
CitSO	LOC18039929	protein PIN-LIKES 6	923	1031	914	971	967	1063	1240	1183	1135	1195	1129	1264	1245	1209	1462
CitERF13	LOC18047942	ethylene-responsive transcription factor 13	2945	2938	2494	1519	1286	1371	1745	1723	1431	1337	1897	1738	3334	4044	4648
CitVHA-c4	LOC18041768	V-type proton ATPase 16 kDa proteolipid subunit	1422	1523	1553	1047	1069	1185	1248	1232	1118	1664	1716	1815	1371	1278	1259

Supplementary Table 7. List of genes validated by PCR in this study.

Gene Name Abb.	Gene Id	Sample	DESeq2 Log_2 FC	ddC_T
CitPH5.1	LOC18035736	S1	-2.74	-1.28
		S2	-2.17	-1.27
		S3	-1.51	-1.65
		ICH	1.33	3.37
CitPH5.2	LOC18035739	S1	-3.45	-4.50
		S2	-0.72	-1.49
		S3	-0.58	-0.81
		ICH	-0.04	-0.53
CitPH1	LOC18037376	S1	-3.11	-2.62
		S2	-0.98	-1.20
		S3	-0.70	-0.02
		ICH	-0.16	-1.78
CitAN1	LOC18047507	S1	-3.59	-4.26
		S2	-1.05	-1.54
		S3	-0.92	-0.37
		ICH	-1.11	-4.64
CitAN11	LOC18032473	S1	-0.16	0.26
		S2	0.48	-0.12
		S3	0.28	0.48
		ICH	0.47	-1.15
CitPH3	LOC18038669	S1	-1.50	-0.41
		S2	-0.72	0.01
		ICH	-0.55	-0.88
CitPH4	LOC18053295	S1	-3.30	-3.87
		S2	-1.09	-0.52
		ICH	-0.46	-0.51
CitERF13	LOC18047942	S1	-0.98	-0.50
		S2	-0.76	-1.44
ase.1.1	LOC18040524	S1	1.50	2.30
		S2	1.27	0.41

		S3	1.16	1.47
gs.4.1	LOC18046053	S1	0.39	-0.56
		S2	1.05	0.68
		S3	1.12	1.08
		ICH	1.67	3.32
gs.4.2	LOC18052541	S1	0.10	0.58
		S2	0.94	1.05
		ICH	1.46	1.81

Supplementary Table 2. DEGs in the pulp of developing fruitlets of ICH versus SCM ($Log_2FC=0.58$ and $\alpha=0.05$).

Supplementary Table 3. DEGs in the pulp of developing fruitlets of ICH as related to SCM ($Log_2FC=1$ and $\alpha=0.01$). DEGs included in the KEGG pathways (Supplementary Figure 3) and genes that did not reach 100 reads in at least one of the replicates, were removed.

Supplementary Table 5. DEGs in the pulp of developing fruitlets of S1, S2 and S3 segregants as realted to SCM. Pairwise comparisons of each segregant, S1, S2 and S3, versus SCM ($Log_2FC=0.58$ and $\alpha=0.05$) were first performed and common DEGs in the three segregants were included in this table. Genes that did not reach at least 100 reads in one of the biological replicates were filtered out. DEGs were grouped into the categories defined in Supplementary Table 3.

They are available in the online version at https://doi.org/10.3389/fpls.2022.982683

DISCUSSION

Discussion

For many centuries, the origin and domestication of citrus has constituted an absolute mystery only approached through a mythological perspective. In recent years, the advent of genomics is providing solid and original information impacting several disciplines related to the taxonomy, the origin, and the evolution and domestication of citrus. Nevertheless, citrus domestication is an unknown issue that we still understand in a very partial and imprecise way. Based on research carried out during the last 20 years (see for instance, Wu et al., 2014; Wang et al., 2017*b*; Wang et al., 2018; Wu et al., 2018), Wu et al., (2021) has prompted the idea that this process is mainly characterized by three major events: emergence of apomixis in the genus *Citrus*, intra- and interspecific introgressions in the mandarin genome, and generation of new variability through both somatic mutations in asexually reproduced varieties and sexual crosses in facultative and non-obligated apomictic genotypes. The work presented in this memory provides original insights into these events.

According to this viewpoint, a fundamental fact in citrus domestication was the acquisition of apomictic behavior by the ancestral citrus species, a form of asexual reproduction that gives rise to polyembryonic seeds, leading to the generation of offspring genetically identical to the maternal plant (Underwood and Mercier, 2022). It has been recently reported that the basic taxon *C. reticulata* is composed of two mandarin subspecies that diversified approximately 1.5 million years ago. One of these groups is composed of those varieties popularly identified as *common mandarins*, while in the other group, represented by *mangshanyeju*, are included other chinese related wild mandarin populations. It is worth to mention that their most recent common ancestor diverged from *C. ryukyuensis* 2.2-2.8 Mya ago (Wu et al., 2021).

During the early Pleistocene, an insertion of a MITE DNA transposon in the promoter of the CitRKD1 gene conferring apomictic behavior (Shimada et al., 2018; Wang et al., 2017*b*), occurred in the *mangshanyeju* genome. These two groups of mandarins interbred and apomixis was first extended through intraspecific introgression into the mandarin pool, and thereafter into several other citrus. Today, the presence of this transposon insertion and the apomictic behavior is found in practically all kind of varieties of edible citrus including mandarins and oranges (Figure 3).

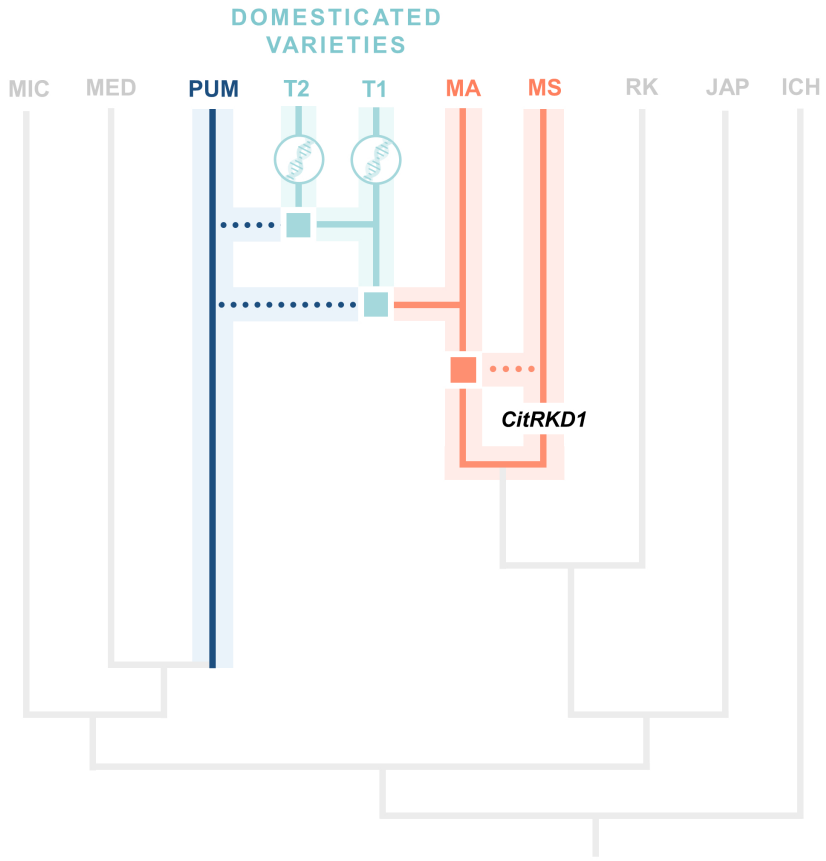


Figure 3. Three major events characterize the process of citrus domestication: the presence of the apomictic behavior represented by the emergence of defective gene *CitRKD1*, the occurrence of pummelo introgression in the mandarin genome, represented by dotted dark blue lines, and the generation of new variability through somatic mutations, represented by a double helix. MIC, *C. micrantha*; MED, *C. medica*; PUM, *C. maxima*; T1, Older mandarins; T2, Modern mandarins; MA, *Common mandarins*, MS, *Mangshanyeju*; RK, *C. ryukyuensis*; JAP, *F. japonica*; ICH, *C. ichangensis*. Adapted from Wu et al., (2018, 2021).

Once the desirable characteristics were acquired, the emergence of apomixis meant in agronomical terms, the possibility of select and harvest fruits with these traits, since the apomictic behavior allows to preserve in time the maternal phenotype. With the development of grafting, that was known and practiced in China more than 2,500 years ago (Melnyk and Meyerowitz, 2015), apomixis was no longer necessary to conserve the desire phenotype. This innovation opened the range of possibilities to other domesticated traits, including those derived from alternative asexual reproductive processes such as parthenocarpy, self-incompatibility or sterility that generate appreciated seedless fruits.

As related to the varieties used in this work, the acidic self-compatible mandarin, Sun Chu Sha Kat (SCM) and the self-incompatible W. Murcott (WMU), contain the defective CitRKD1 gene (Wu et al., 2021) and show nucellar polyembryony. On the other hand, the wild Ichang papeda (ICH) and Clementine (CLM) are both monoembryonic and develop sexual seeds, although the first one is self-compatible and Clementine, in contrast, is self-incompatible.

The RNA-seq data indicate that major differences in the regulation of gene expression in the fruitlet flesh of wild papeda and acidic SCM mandarin, affected basic processes of plant growth and development, including pivotal central pathways of the primary and secondary metabolism (Chapter 2: Figure 1, Supplementary Table 2 and Supplementary Table 3). For instance, main biosynthetic routes of secondary metabolism including alkaloids, glucosinolates, cyanogenic glycosides, phytoalexins and other groups of chemical defenses are upregulated in the wild variety in comparison with SCM (Figure 4). In other crops, domestication is also accompanied by a reduction of secondary metabolism concomitant with an increase in growth and development, especially reproductive growth (Gonzalez-Ibeas et al., 2021*b*; Rao et al., 2020; Herms and Mattson, 1992; Rosenthal and Dirzo, 1997; Meyer et al., 2012).

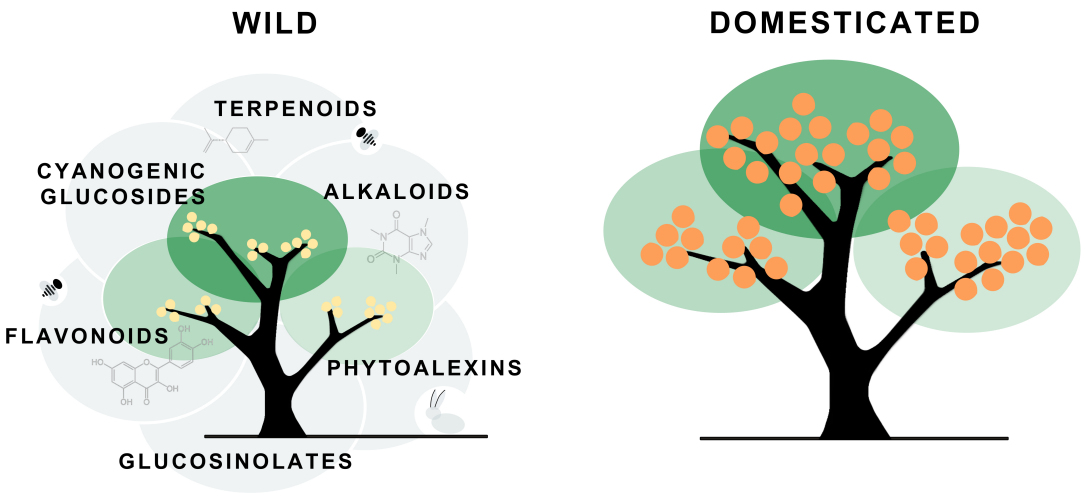


Figure 4. Major differences in gene expression in wild and domesticated citrus are related to the production of secondary metabolites in non-cultivated species and to the stimulation of growth processes in domesticated varieties. In these citrus, the decline in secondary metabolism mainly involves reduction of distasteful chemical defenses against herbivorous insects, a circumstance that results in increased palatability.

Differences between SCM and the palatable mandarins were, on the other hand, restricted to a scarce number of DEGs, most of them implicated in palatability (Chapter 2: Figure 4 and Supplementary Table 5). Taken together, these observations strongly suggest that SCM, although develops acidic fruits, is probably an early product of the domestication processes of citrus. In this regard, Tanaka (1954), reports that SCM certainly was cultivated in certain areas of Japan. SCM contains a pure mandarin genome, albeit in strict sense is a genetic admixture since as mentioned above, carries *mangshanyeju* introgression. This observation is interesting since current commercial mandarins unequivocally show signatures of interspecific pummelo (*C. maxima*) introgression, a circumstance that in addition to apomixis was similarly spread among other edible type of citrus. Consequently, domesticated mandarins, oranges and grapefruits are genetic admixtures containing fragments of *mangshanyeju*, pummelo and *common mandarin* genomes (Wu et al., 2014, 2018, 2021). This circumstance clearly indicates that these three types of citrus are the foundational species of the domesticated, edible citrus. The data also imply that *mangshanyeju* introgression predated pummelo hybridization. The contribution of pummelo stretches appear to be important for the development of edible varieties, since the results of the differential allelic expression study presented in Chapter 2: Figure 4 and Supplementary Figure 8 are characterized by the downregulation of the mandarin alleles and the upregulation of the pummelo ones, suggesting that their incorporations were selected and preserved over time. Therefore, mandarin domestication was partially based on the selection and substitution of mandarin alleles by pummelo genes. Pummelo introgressions, however, do not appear to be related to the accumulation of citric acid, although the analysis detected a few targets that in principle might be contributors to other relevant traits, such as sweetness (Chapter 2: Figure 4, Supplementary Table 5, Supplementary Figure 8).

The combination of these two phenomena, apomixis and hybridization, leads inexorably to reticulate evolution (Figure 3) (Hamston et al., 2018), where apomixis results in the *freezing* of the hybrid genotype, while variability, if present, is provided by somatic mutations or new sexual crosses in those varieties that scape to apomixis. The huge number of current varieties of mandarins and oranges, indicates that the later stages of citrus domestication, the third relevant event in this process, are typified by the thousands of somatic mutations and myriads of crosses that have occurred in asexual

and sexual genotypes, respectively, to produce the elite varieties that are currently displayed in our supermarkets.

The results of this work also show that the most important target of this third stage of the domestication process was the increase in palatability, which is supported by the differential regulation of gene expression observed between the antique acidic SCM and the palatable mandarins derived from modern elite varieties (Chapter 2: Supplementary Table 5). Thus, the group of genes involved in the regulation of acidity were indeed downregulated in the edible varieties (Chapter 2: Figure 2 and Figure 3), as were those implicated in the catalysis of sucrose, while certain cytosolic sugar transporters were upregulated. Other upregulated DEGs that may affect palatability and taste were these controlling last steps in the synthesis of methyl anthranilate and furaneol. The first one is a substance of pleasant aroma, involved in the fragrance of Concord grapes (*Vitis labrusca* L., 1753) (Wang and De Luca, 2005), that has been used in flavoring foods as mandarin candies or soft drinks (cites in Luo et al., 2019; Lee et al., 2019). Furaneol, on the other hand, is a key flavor compound in strawberries (*Fragaria* × *ananassa* (Weston) Duchesne ex Rozier) (Raab et al., 2006). It is worth to note that expression of methanol O-anthraniloyltransferase was barely detectable in the inedible papeda, but its expression was relatively high in SCM and even higher in the three segregants. Consistently, pivotal genes implicated in the synthesis of secondary metabolites, that generally are unpalatable compounds, included in the groups of terpenoids, alkaloids, glucosinolates, and polyphenols were downregulated. It is also interesting to point out that palatable mandarins showed downregulation of one of the main steps in the synthesis of scopolamine, which is a potent psychoactive drug. On the contrary, they show upregulation of genes involved in the production of serotonin and dopamine, which act as hormones and are known as the *drugs of happiness*. This last observation may seem suggestive or evocative, although at the moment its relevance is not understand because, serotonin for example, is not easily accessible to the central nervous system due to blood brain barrier (Shabbir et al., 2013). Finally, results also suggest that cell wall stiffening of the juice sacs in the fruitlet pulp of the palatable mandarins is reduced, a characteristic that may be associated with a higher degree of palatability.

Most of these changes in gene regulation implicated in the domestication process were likely related to the occurrence of somatic mutations, since in the absence of sexual

reproduction, variability and diversity are exclusively generated through mutations. Furthermore, with the discovery and use of new practices of asexual reproduction, a huge amount of seedless citrus types, including parthenocarpic, sterile and self-incompatible mandarins and oranges, were very likely selected during the last thousands of years. For example, Wang et al., (2021) have reported in a comprehensive study the somatic mutations detected in the orange genome, including SNPs, structural variations and transposable elements. In previous works, we have shown that somatic mutations that involve relatively large fragments, may generate phenotypes that are easily distinguishable of the original variety. The Arrufatina mandarin, for instance, is an early Clementine variety containing a huge deletion of 2 Mb in chromosome 3 (Terol et al., 2015), that appears to be responsible of its earliness behavior (Terol et al., 2019). This structural variation, as many others, is due to the activity of transposable elements, that are very dynamic in citrus (Borredá et al., 2019), and can be inserted in the regulatory regions of relevant genes modifying their expression (Butelli et al., 2012; Shimada et al., 2018; Strazzer et al., 2019). Similarly, SNPs may affect certain or specific attributes of interest, but generally do not drastically alter other desirable phenotypes, a characteristic compatible for instance, with the observation that many regulatory genes of acidity in the three palatable mandarins, are downregulated but are still functional. The downregulation of main routes involved in the secondary metabolism observed in SCM also seems to be compatible with the idea that point mutations may have significant effects when produced over extended periods of time. These observations are consistent with the suggestion that favorable point mutations affecting a certain desirable trait, may have been continuously selected during the domestication process. An illustrative example of this continuous process of selection has been recently proposed for seed shattering, one of the most important domestication traits in rice, which is regulated by two genes. One of these genes appears to be the real domestication gene of the non-shattering trait, since it was the target of an early selection that occurred more than 8,000 years ago, while the selection for the other gene proceeded thereafter over a period spanning three millennia (Zheng et al., 2016). According to Izawa, (2022), the first gene is a real domestication gene, while the second one appears to be associated with a continuous post-domestication improvement. Furthermore, Ishikawaa et al., (2022) in a recent paper, have suggested a stepwise route for the earliest phase of rice domestication, including the sequential recruitment of several genes that together led to the loss of shattering.

In this work, we have also characterized in Clementine, the dynamics of single-nucleotide mosaicism, a process that presumably has played key roles driving citrus domestication. Briefly, our results indicate that the generation and distribution of somatic variants are stochastic processes. The distribution of somatic mutations, for instance, was quite uniform along all Clementine chromosomes. The number of variants found in a given chromosome was equally distributed between homozygous mandarin regions and the stretches characterized by pummelo introgression. Similarly, the allocation of mutations also appears to be uniform in areas with high and low heterozygosity. The data show that regions with low gene prevalence, such as pericentromeric and transposon areas, contain the same number of nucleotide substitutions found in the regions characterized by a high presence of genes (Chapter 1: Figure 1). Monroe et al., (2022), however, in an exhaustive study in Arabidopsis, have found that mutations occur more frequently in those areas containing lower number of genes and that the regions associated with the epigenome show a significant reduction of deleterious mutations. In perennials, mutations accumulate in sectorized areas in a hierarchical way (Schmid-Siegert et al., 2017), that in citrus is determined by the typical sympodial branching model. This implies both that the distribution of somatic mutations follows an iterative process according to the branching pattern, and that variants accumulate over time, so that younger tissues have higher mutation rates. In this scenario, a citrus tree can be considered as a genetic mosaic composed of different genomes arranged in sectors (Chapter 1: Figure 4, Figure 5 and Table 1).

Estimate of mutation rates in Clementine was 4.4×10^{-10} bp^{-1} yr^{-1}, a rate that is remarkably similar to that reported for other perennial trees, for example, peaches (*Prunus persica* (L.) Batsch) (Wang et al., 2019). Our estimates also indicate that the total number of variants present in a commercial Clementine tree with the characteristics of the one we have analyzed, ranges between 1,500 and 5,000, which on average implies, the presence of 1 mutation per developing sprout. Basically, these results suggest that the mutations rates that occur in somatic tissues of citrus are relatively high and, therefore, compatible with the idea that they may have certainly played an important role in citrus domestication, as suggested in this work for example, in the generation of palatable mandarins. The genetic mosaic model, hierarchized by sectors according to iterative ramifications, constitutes the basis of the asexual selection process that still supports modern citriculture. Most of the current citrus varieties that

populate our markets have undergone a manual selection process by citrus growers, who upon detecting new characteristics in a certain branch, fixed these novel attributes through grafting and subsequent propagation. This is the same procedure that has been practiced by farmers since time immemorial in China and that progressively led to the domestication of citrus.

CONCLUSIONS

Conclusions

1. The distribution of single-nucleotide variants in Clementine (*C. clementina*) trees follows an iterative process determined by the branching pattern. The hierarchical nature of the process arranges mutations in sectors according to the sympodial growth habit of this species.

2. Estimate of mutation rate in Clementine was 4.4×10^{-10} bp^{-1} yr^{-1}, and as observed in younger branches, variants accumulated with time. The analysis also revealed that these trees carry between 1,500 and 5,000 mutations and that on average, each leaf flush generates approximately 1 new variant.

3. Taken together, the previous observations indicate that Clementine trees are mosaics genetically composed of different genomes sectored by areas.

4. The RNA-seq analyses of the pulp of developing fruitlets of inedible papeda and acidic mandarin, suggest that overall, domestication stimulates basal growth at the expense of the reduction of chemical defenses, including alkaloids, terpenoids, phenylpropanoids, flavonoids, glucosinolates and cyanogenic glucosides. Those changes led to major differences in organoleptic properties since the reduction of unpleasant secondary metabolites increases palatability.

5. The comparison between palatable and acidic mandarins identified a small set of downregulated genes in the edible mandarins, mostly involved in the accumulation of citric acid, and to a lesser extent, in the synthesis of secondary metabolites and cell wall components.

6. Based on these results, it is proposed that citrus domestication involved:

   a) a first transition phase, from inedible papeda to sour mandarin, characterized by a drastic and general reprogramming of gene regulation of central pathways of primary and secondary metabolism;

   b) a second phase, in which edible attributes of modern mandarins, such as pleasant aroma and flavor, and especially acidity, were refined mainly through specific progressive changes. Estimation of somatic mutation rates are compatible with the idea that they played important roles during these later steps of citrus domestication.

REFERENCES

References

Alquézar, B., Zacarías, L., & Rodrigo, M. J. (2009). Molecular and functional characterization of a novel chromoplast-specific lycopene beta-cyclase from Citrus and its relation to lycopene accumulation. *Journal of experimental botany*, 60(6), 1783–1797. https://doi.org/10.1093/jxb/erp048

Borredá, C., Pérez-Román, E., Ibanez, V., Terol, J., & Talon, M. (2019). Reprogramming of Retrotransposon Activity during Speciation of the Genus Citrus. *Genome biology and evolution*, 11(12), 3478–3495. https://doi.org/10.1093/gbe/evz246

Borredá, C., Perez-Roman, E., Talon, M., & Terol, J. (2022). Comparative transcriptomics of wild and commercial Citrus during early ripening reveals how domestication shaped fruit gene expression. *BMC plant biology*, 22(1), 123. https://doi.org/10.1186/s12870-022-03509-9

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., & Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *The Plant cell*, 24(3), 1242–1255. https://doi.org/10.1105/tpc.111.095232

Butelli, E., Garcia-Lor, A., Licciardello, C., Las Casas, G., Hill, L., Recupero, G. R., Keremane, M. L., Ramadugu, C., Krueger, R., Xu, Q., Deng, X., Fanciullino, A. L., Froelicher, Y., Navarro, L., & Martin, C. (2017). Changes in Anthocyanin Production during Domestication of Citrus. *Plant physiology*, 173(4), 2225–2242. https://doi.org/10.1104/pp.16.01701

Butelli, E., Licciardello, C., Ramadugu, C., Durand-Hulak, M., Celant, A., Recupero, G. R., Froelicher, Y., & Martin, C. (2019). Noemi controls production of flavonoid pigments and fruit acidity and illustrates the domestication routes of modern citrus varieties. *Current biology: CB*, 29(1), 158-164. https://doi.org/10.1016/j.cub.2018.11.040

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., & Dopazo, J. (2015). A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus Citrus. *Molecular biology and evolution*, 32(8), 2015–2035. https://doi.org/10.1093/molbev/msv082

Caruso, M., Casas, G., Scaglione, D., Gattolin, S., Rossini, L., Distefano, G., Cattonaro, F., Catara, A., Licciardello, G., Morgante, M., & Licciardello, C. (2018). Detection of natural and induced mutations from next generation sequencing data in sweet orange bud sports. *Acta Horticulturae*, 119-124. https://doi.org/10.17660/ActaHortic.2018.12.15.

Cercós, M., Soler, G., Iglesias, D. J., Gadea, J., Forment, J., & Talón, M. (2006). Global analysis of gene expression during development and ripening of citrus fruit flesh. A proposed mechanism for citric Acid utilization. *Plant molecular biology*, 62(4-5), 513–527. https://doi.org/10.1007/s11103-006-9037-

De La Torre, A. R., Li, Z., Van de Peer, Y., & Ingvarsson, P. K. (2017). Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. *Molecular Biology and Evolution*, 34(6), 1363–1377. https://doi.org/10.1093/molbev/msx069

EUROSTAT Comext, 2021, https://ec.europa.eu/info/food-farming-fisheries/farming/facts-and-figures/markets/overviews/market-observatories/fruit-and-vegetables/citrus-fruit-statistics_en

FAO.FAOSTAT statistical database. License: CC BY-NC-SA 3.0 IGO. Extracted from: https://www.fao.org/faostat/en/#data. Data of Access: 02-07-2022.

Fu, S., Shao, J., Paul, C., Zhou, C., & Hartung, J. S. (2017). Transcriptional analysis of sweet orange trees co-infected with 'Candidatus Liberibacter asiaticus' and mild or severe strains of Citrus tristeza virus. *BMC Genomics*, 18(1), 837. https://doi.org/10.1186/s12864-017-4174-8

Gabur, I., Chawla, H. S., Snowdon, R. J., & Parkin, I. (2019). Connecting genome structural variation with complex traits in crop plants. *TAG. Theoretical and applied genetics*, 132(3), 733–750. https://doi.org/10.1007/s00122-018-3233-0

Gonzalez-Ibeas, D., Ibanez, V., Perez-Roman, E., Borredá, C., Terol, J., & Talon, M. (2021*a*). Shaping the biology of citrus: I. Genomic determinants of evolution. *The Plant genome*, 14(3), e20133. https://doi.org/10.1002/tpg2.20104

Gonzalez-Ibeas, D., Ibanez, V., Perez-Roman, E., Borredá, C., Terol, J., & Talon, M. (2021*b*). Shaping the biology of citrus: II. Genomic determinants of domestication. *The plant Genome*, 14(3), e20133. https://doi.org/10.1002/tpg2.20133

Hamston, T. J., de Vere, N., King, R. A., Pellicer, J., Fay, M. F., Cresswell, J. E., & Stevens, J. R. (2018). Apomixis and Hybridization Drives Reticulate Evolution and Phyletic Differentiation in Sorbus L.: Implications for Conservation. *Frontiers in plant science*, 9, 1796. https://doi.org/10.3389/fpls.2018.01796

Hanlon, V., Otto, S. P., & Aitken, S. N. (2019). Somatic mutations substantially increase the per-generation mutation rate in the conifer Picea sitchensis. Evolution letters, 3(4), 348–358. https://doi.org/10.1002/evl3.121

Herms, D. A., & Mattson, W. J. (1992). The dilemma of plants: to grow or defend. *The quarterly review of biology*, 67(3), 283-335.

Hussain, S.B., Shi, C.Y., Guo, L.X., Kamran, H., Sadka, A., & Liu, Y. (2017). Recent Advances in the Regulation of Citric Acid Metabolism in Citrus Fruit. *Critical Reviews in Plant Sciences*, 36. 1-16. 10.1080/07352689.2017.1402850.

Hussain, S. B., Guo, L. X., Shi, C. Y., Khan, M. A., Bai, Y. X., Du, W., & Liu, Y. Z. (2020). Assessment of sugar and sugar accumulation-related gene expression profiles reveal new insight into the formation of low sugar accumulation trait in a sweet orange (Citrus sinensis) bud mutant. *Molecular biology reports*, 47(4), 2781–2791. https://doi.org/10.1007/s11033-020-05387-6

Ishikawa, R., Castillo, C. C., Htun, T. M., Numaguchi, K., Inoue, K., Oka, Y., Ogasawara, M., Sugiyama, S., Takama, N., Orn, C., Inoue, C., Nonomura, K. I., Allaby, R., Fuller, D. Q., & Ishii, T. (2022). A stepwise route to domesticate rice by controlling seed shattering and panicle shape. *Proceedings of the National Academy of Sciences of the United States of America*, 119(26), e2121692119. https://doi.org/10.1073/pnas.2121692119

Izawa T. (2022). Reloading DNA History in Rice Domestication. *Plant Cell Physiology*. pcac073. Advance online publication. https://doi.org/10.1093/pcp/pcac073

Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., Davey, J. W., & Jiggins, C. D. (2015). Estimation of the spontaneous mutation rate in Heliconius melpomene. *Molecular biology and evolution*, 32(1), 239–243. https://doi.org/10.1093/molbev/msu302

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. T., Thorsteinsdottir, U., … Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, 488(7412), 471–475. https://doi.org/10.1038/nature11396

Lee, H. L., Kim, S. Y., Kim, E. J., Han, D. Y., Kim, B. G., & Ahn, J. H. (2019). Synthesis of Methylated Anthranilate Derivatives Using Engineered Strains of Escherichia coli. *Journal of microbiology and biotechnology*, 29(6), 839–844. https://doi.org/10.4014/jmb.1904.04022

Liu, X., Hu, X. M., Jin, L. F., Shi, C. Y., Liu, Y. Z., & Peng, S. A. (2014). Identification and transcript analysis of two glutamate decarboxylase genes, CsGAD1 and CsGAD2, reveal the strong relationship between CsGAD1 and citrate utilization in citrus fruit. *Molecular biology reports*, 41(9), 6253–6262. https://doi.org/10.1007/s11033-014-3506-x

Luo, Z. W., Cho, J. S., & Lee, S. Y. (2019). Microbial production of methyl anthranilate, a grape flavor compound. *Proceedings of the National Academy of Sciences of the United States of America*, 116(22), 10749–10756. https://doi.org/10.1073/pnas.1903875116

Lv, X., Zhao, S., Ning, Z., Zeng, H., Shu, Y., Tao, O., Xiao, C., Lu, C., & Liu, Y. (2015). Citrus fruits as a treasure trove of active natural metabolites that potentially provide benefits for human health. *Chemistry Central journal*, 9, 68. https://doi.org/10.1186/s13065-015-0145-9

Lynch M. (2010). Evolution of the mutation rate. *Trends Genetics*. 26(8), 345–352. https://doi.org/10.1016/j.tig.2010.05.003

Maddi, T., Pérez-Román, E., Maiza-Benabdesselam, F., Khettal, B., Talon, M., & Ibanez-Gonzalez, V. (2018). New Citrus chloroplast haplotypes revealed by molecular markers using Algerian and Spanish accessions. *Genetic resources*

and crop evolution, 65(8), 2199-2214. https://doi.org/10.1007/s10722-018-0685-7

MAPA. 2022. Superficies y producciones anuales de cultivo. Madrid, Spain: Ministerio de Agricultura, Pesca y Alimentación

Melnyk, C. W., & Meyerowitz, E. M. (2015). Plant grafting. *Current biology*, 25(5), R183–R188. https://doi.org/10.1016/j.cub.2015.01.029

Meyer, R. S., DuVal, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New phytologist*, 196(1), 29–48. https://doi.org/10.1111/j.1469-8137.2012.04253.x

Miller, A. J., & Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *American journal of botany*, 98(9), 1389–1414. https://doi.org/10.3732/ajb.1000522

Monroe, J., Srikant, T., Carbonell-Bejerano, P., Becker, C., Lensink, M., Exposito-Alonso, M., Klein, M., Hildebrandt, J., Neumann, M., Kliebenstein, D., Weng, M. L., Imbert, E., Ågren, J., Rutter, M. T., Fenster, C. B., & Weigel, D. (2022). Mutation bias reflects natural selection in Arabidopsis thaliana. *Nature*, 602(7895), 101-105. https://doi.org/10.1038/s41586-021-04269-6

National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2022 July 05]. Available from: https://www.ncbi.nlm.nih.gov/

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D., & Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science*, 327(5961), 92–94. https://doi.org/10.1126/science.1180677

Primo-Capella, A., Forner-Giner, M. Á., Martínez-Cuenca, M. R., & Terol, J. (2022). Comparative transcriptomic analyses of citrus cold-resistant vs. sensitive rootstocks might suggest a relevant role of ABA signaling in triggering cold scion adaption. *BMC plant biology*, 22(1), 209. https://doi.org/10.1186/s12870-022-03578-w

Purugganan M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication. *Current biology*, 29(14), R705–R714. https://doi.org/10.1016/j.cub.2019.05.053

Raab, T., López-Ráez, J. A., Klein, D., Caballero, J. L., Moyano, E., Schwab, W., & Muñoz-Blanco, J. (2006). FaQR, required for the biosynthesis of the strawberry flavor compound 4-hydroxy-2,5-dimethyl-3(2H)-furanone, encodes an enone oxidoreductase. *The Plant cell*, 18(4), 1023–1037. https://doi.org/10.1105/tpc.105.039784

Ramírez-Pool, J. A., Xoconostle-Cázares, B., Calderón-Pérez, B., Ibarra-Laclette, E., Villafán, E., Lira-Carmona, R., & Ruiz-Medrano, R. (2022). Transcriptomic Analysis of the Host Response to Mild and Severe CTV Strains in Naturally

Infected Citrus sinensis Orchards. *International journal of molecular sciences*, 23(5), 2435. https://doi.org/10.3390/ijms23052435

Rao, M. J., Zuo, H., & Xu, Q. (2020). Genomic insights into citrus domestication and its important agronomic traits. *Plant Communications*, 2(1), 100138. https://doi.org/10.1016/j.xplc.2020.100138

Rosenthal, J.P., & Dirzo, R. (1997). Effects of life history, domestication and agronomic selection on plant defence against insects: Evidence from maizes and wild relatives. *Evolutionary Ecology*. 11, 337–355. https://doi.org/10.1023/A:1018420504439

Schmid-Siegert, E., Sarkar, N., Iseli, C., Calderon, S., Gouhier-Darimont, C., Chrast, J., Cattaneo, P., Schütz, F., Farinelli, L., Pagni, M., Schneider, M., Voumard, J., Jaboyedoff, M., Fankhauser, C., Hardtke, C. S., Keller, L., Pannell, J. R., Reymond, A., Robinson-Rechavi, M., Xenarios, I., … Reymond, P. (2017). Low number of fixed somatic mutations in a long-lived oak tree. *Nature plants*, 3(12), 926–929. https://doi.org/10.1038/s41477-017-0066-9

Schrider, D. R., Houle, D., Lynch, M., & Hahn, M. W. (2013). Rates and genomic consequences of spontaneous mutational events in Drosophila melanogaster. *Genetics*. 194(4), 937–954. https://doi.org/10.1534/genetics.113.151670

Shabbir, F., Patel, A., Mattison, C., Bose, S., Krishnamohan, R., Sweeney, E., Sandhu, S., Nel, W., Rais, A., Sandhu, R., Ngu, N., & Sharma, S. (2013). Effect of diet on serotonergic neurotransmission in depression. *Neurochemistry international*, 62(3), 324–329. https://doi.org/10.1016/j.neuint.2012.12.014

Sheng, L., Shen, D., Luo, Y., Sun, X., Wang, J., Luo, Zeng, Y., Xu, J., Deng, X., & Cheng, Y. (2017). Exogenous γ-aminobutyric acid treatment affects citrate and amino acid accumulation to improve fruit quality and storage performance of postharvest citrus fruit. *Food chemistry*. 216, 138–145. https://doi.org/10.1016/j.foodchem.2016.08.024

Shimada, T., Nakano, R., Shulaev, V., Sadka, A., & Blumwald, E. (2006). Vacuolar citrate/H+ symporter of citrus juice cells. *Planta*, 224(2), 472–480. https://doi.org/10.1007/s00425-006-0223-2

Shimada, T., Endo, T., Fujii, H., Nakano, M., Sugiyama, A., Daido, G., Ohta, S., Yoshioka, T., & Omura, M. (2018). MITE insertion-dependent expression of CitRKD1 with a RWP-RK domain regulates somatic embryogenesis in citrus nucellar tissues. *BMC Plant Biology*, 18(1), 166. https://doi.org/10.1186/s12870-018-1369-3

Smith, S. A., & Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science*. 322(5898), 86–89. https://doi.org/10.1126/science.1163197

Spreen, T.H., Gao, Z., Fernandes, W. Jr., & Zansler, M.L., (2020). "Global economics and marketing of citrus products" in The Genus Citrus, ed. Talon, M., Caruso, M. and Gmitter F.G., (Woodhead Publishing), 471–493.

Strazzer, P., Spelt, C. E., Li, S., Bliek, M., Federici, C. T., Roose, M. L., Koes, R., & Quattrocchio, F. M. (2019). Hyperacidification of Citrus fruits by a vacuolar proton-pumping P-ATPase complex. *Nature communications*, 10(1), 744. https://doi.org/10.1038/s41467-019-08516-3

Swingle W.T., & Reece P.C. (1967). "History, world distribution, botany, and varieties" in The Citrus Industry, revised 2nd, vol 1, ed. Reuther W., Webber H.J. and Batchelor L.D. (Berkeley: University of California Press), 190–430.

Tadeo, F., Terol, J., Rodrigo, M.J., Licciardello, C., & Sadka, A. (2020). "Fruit growth and development" in The Genus Citrus, ed. Talon, M., Caruso, M. and Gmitter F.G., (Woodhead Publishing), 245-269.

Tanaka, T. (1954) Species problem in Citrus. Japanese Society for Promotion of Science.

Terol, J., Soler, G., Talon, M., & Cercos, M. (2010). The aconitate hydratase family from Citrus. *BMC plant biology*, 10, 222. https://doi.org/10.1186/1471-2229-10-222

Terol, J., Ibañez, V., Carbonell, J., Alonso, R., Estornell, L. H., Licciardello, C., Gut, I. G., Dopazo, J., & Talon, M (2015). Involvement of a citrus meiotic recombination TTC-repeat motif in the formation of gross deletions generated by ionizing radiation and MULE activation. *BMC Genomics*, 16(1), 69. https://doi.org/10.1186/s12864-015-1280-3

Terol, J., Tadeo, F., Ventimilla, D., & Talon, M. (2016). An RNA-Seq-based reference transcriptome for Citrus. *Plant Biotechnology Journal*, 14(3), 938–950. https://doi.org/10.1111/pbi.12447

Terol, J., Nueda, M. J., Ventimilla, D., Tadeo, F., & Talon, M. (2019). Transcriptomic analysis of Citrus clementina mandarin fruits maturation reveals a MADS-box transcription factor that might be involved in the regulation of earliness. *BMC Plant Biology*, 19(1), 47. https://doi.org/10.1186/s12870-019-1651-z

Uchimura, A., Higuchi, M., Minakuchi, Y., Ohno, M., Toyoda, A., Fujiyama, A., Miura, I., Wakana, S., Nishino, J., & Yagi, T. (2015). Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Research*, 25(8), 1125–1134. https://doi.org/10.1101/gr.186148.114

Underwood, C. J., & Mercier, R. (2022). Engineering Apomixis: Clonal Seeds Approaching the Fields. *Annual review of plant biology*, 73, 201–225. https://doi.org/10.1146/annurev-arplant-102720-013958

Venn, O., Turner, I., Mathieson, I., de Groot, N., Bontrop, R., & McVean, G. (2014). Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. *Science*. 344(6189), 1272–1275. https://doi.org/10.1126/science.344.6189.1272

Wang, J., & De Luca, V. (2005). The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. *Plant Journal*, 44(4), 606–619. https://doi.org/10.1111/j.1365-313X.2005.02552.x

Wang, J. H., Liu, J. J., Chen, K. L., Li, H. W., He, J., Guan, B., & He, L. (2017*a*). Comparative transcriptome and proteome profiling of two Citrus sinensis cultivars during fruit development and ripening. *BMC Genomics*. 18(1), 984. https://doi.org/10.1186/s12864-017-4366-2

Wang, L., He, F., Huang, Y., He, J., Yang, S., Zeng, J., Deng, C., Jiang, X., Fang, Y., Wen, S., Xu, R., Yu, H., Yang, X., Zhong, G., Chen, C., Yan, X., Zhou, C., Zhang, H., Xie, Z., Larkin, R. M., … Xu, Q. (2018). Genome of wild mandarin and domestication history of mandarin. *Molecular plant*, 11(8), 1024-1037. https://doi.org/10.1016/j.molp.2018.06.001

Wang, L., Ji, Y., Hu, Y., Hu, H., Jia, X., Jiang, M., Zhang, X., Zhao, L., Zhang, Y., Jia, Y., Qin, C., Yu, L., Huang, J., Yang, S., Hurst, L. D., & Tian, D. (2019). The architecture of intra-organism mutation rate variation in plants. *PLoS biology*, 17(4), e3000191. https://doi.org/10.1371/journal.pbio.3000191

Wang, L., Huang, Y., Liu, Z., He, J., Jiang, X., He, F. Lu, Z., Yang, S., Chen, P., Yu, H., Zeng, B., Ke, L., Xie, Z., Larkin, R. M., Jiang, D., Ming, R., Buckler, E. S., Deng, X., & Xu, Q. (2021). Somatic variations led to the selection of acidic and acidless orange cultivars. *Nature plants*, 7(7), 954–965. https://doi.org/10.1038/s41477-021-00941-x

Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., Wu, G., Tian, S., Chen, C., Liu, Y., Yu, H., Yang, X., Lan, H., Wang, N., Wang, L., Xu, J., Jiang, X., Xie, Z., Tan, M., Larkin, R. M., … Xu, Q. (2017*b*). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature genetics*, 49, 765–772. https://doi.org/10.1038/ng.3839

Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., P errier, X., Ruiz, M., Scalabrin, S., Terol, J., Takita, M. A., Labadie, K., Poulain, J., Couloux, A., Jabbari, K., Cattonaro, F., Del Fabbro, C., Pinosio, S., Zuccolo, A., Chapman, J., … Rokhsar, D. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature biotechnology*, 32 (7):656-62. https://doi.org/10.1038/nbt.2906

Wu, G.A., Terol J., Ibanez, V., Lopez-Garcia, A., Perez-Roman, E., Borredá, C., Domingo, C., Tadeo, F. R., Carbonell-Caballero, J., Alonso, R., Curk, F., Du, D., Ollitrault, P., Roose, M. L., Dopazo, J., Gmitter, F. G., Rokhsar, D. S., & Talon, M. (2018). Genomics of the origin, evolution and domestication of citrus. *Nature*, 544:311–316. https://doi.org/10.1038/nature25447

Wu, G. A., Sugimoto, C., Kinjo, H., Azama, C., Mitsube, F., Talon, M., Gmitter, F. G., Jr, & Rokhsar, D. S. (2021). Diversification of mandarin citrus by hybrid speciation and apomixis. *Nature communications*, 12(1), 1-10. https://doi.org/10.1038/s41467-021-24653-0

Wu, J., Fu, L., & Yi, H. (2016). Genome-Wide Identification of the Transcription Factors Involved in Citrus Fruit Ripening from the Transcriptomes of a Late-Ripening Sweet Orange Mutant and Its Wild Type. *PLoS One*, 11(4), e0154330. https://doi.org/10.1371/journal.pone.0154330

Xie, S., Manchester, S. R., Liu, K., Wang, Y., Sun, B., & Editor: Patrick Herendeen. (2013). Citrus linczangensis sp. n., a Leaf Fossil of Rutaceae from the Late Miocene of Yunnan, China. *International Journal of Plant Sciences*, 174(8), 1201–1207. https://doi.org/10.1086/671796

Xie, Z., Wang, L., Wang, L., Wang, Z., Lu, Z., Tian, D., Yang, S., & Hurst, L. D. (2016). Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. A low rate in woody perennials and a higher mutagenicity in hybrids. *Proceedings. Biological sciences*, 283(1841), 20161016. https://doi.org/10.1098/rspb.2016.1016

Xu, Q., Chen, L. L., Ruan, X., Chen, D., Zhu, A., Chen, C., Bertrand, D., Jiao, W. B., Hao, B. H., Lyon, M. P., Chen, J., Gao, S., Xing, F., Lan, H., Chang, J. W., Ge, X., Lei, Y., Hu, Q., Miao, Y., Wang, L., … Ruan, Y. (2013). The draft genome of sweet orange (Citrus sinensis). *Nature genetics*, 45(1), 59–66. https://doi.org/10.1038/ng.2472

Yang, S., Wang, L., Huang, J., Zhang, X., Yuan, Y., Chen, J. Q., Hurst, L. D., & Tian, D. (2015). Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature*, 523(7561), 463–467. https://doi.org/10.1038/nature14649

Zhang, L., Ma, G., Shirai, Y., Kato, M., Yamawaki, K., Ikoma, Y., & Matsumoto, H. (2012). Expression and functional analysis of two lycopene β-cyclases from citrus fruits. *Planta*, 236(4), 1315–1325. https://doi.org/10.1007/s00425-012-1690-2

Zhang, H., Chen, J., Peng, Z., Shi, M., Liu, X., Wen, H., Jiang, Y., Cheng, Y., Xu, J., & Zhang, H. (2021). Integrated Transcriptomic and Metabolomic analysis reveals a transcriptional regulation network for the biosynthesis of carotenoids and flavonoids in 'Cara cara' navel Orange. *BMC plant biology*, 21(1), 29. https://doi.org/10.1186/s12870-020-02808-3

Zheng, X., Zhu, K., Sun, Q., Zhang, W., Wang, X., Cao, H., Tan, M., Xie, Z., Zeng, Y., Ye, J., Chai, L., Xu, Q., Pan, Z., Xiao, S., Fraser, P. D., & Deng, X. (2019). Natural Variation in CCD4 Promoter Underpins Species-Specific Evolution of Red Coloration in Citrus Peel. *Molecular plant*, 12(9), 1294–1307. https://doi.org/10.1016/j.molp.2019.04.014

Zheng, Y., Crawford, G. W., Jiang, L., & Chen, X. (2016). Rice Domestication Revealed by Reduced Shattering of Archaeological rice from the Lower Yangtze valley. *Scientific reports*, 6, 28136. https://doi.org/10.1038/srep28136

Zhu, C., Zheng, X., Huang, Y., Ye, J., Chen, P., Zhang, C., Zhao, F., Xie, Z., Zhang, S., Wang, N., Li, H., Wang, L., Tang, X., Chai, L., Xu, Q., & Deng, X. (2019). Genome sequencing and CRISPR/Cas9 gene editing of an early flowering Mini-Citrus (Fortunella hindsii). *Plant biotechnology journal*, 17(11), 2199-2210. https://doi.org/10.1111/pbi.13132