Received: 24 March 2021 | Accepted: 2 September 2021

DOI: 10.1002/tpg2.20162

The Plant Genome

ORIGINAL RESEARCH

Single-nucleotide mosaicism in citrus: Estimations of somatic mutation rates and total number of variants

Estela Perez-Roman | **Carles Borredá** | **Antoni López-García Usach** | **Manuel Talon**

Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada, Valencia 46113, Spain

Correspondence
Manuel Talon, Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada 46113, Valencia, Spain
Email: talon_man@gva.es

Assigned to Associate Editor Barbara Blanco-Ulate.

Funding information
Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: RTI2018-097790-R-100; Generalitat Valenciana, Instituto Valenciano de Investigaciones Agrarias, Grant/Award Numbers: 51915, 52002

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.

1 | 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 & 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

Abbreviations: IGV, Integrative Genomics Viewer; SNP, single-nucleotide polymorphism; UV, ultraviolet.

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 & 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; L. 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 & 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 & 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).

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.

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 & 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, L. 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; L. 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.

2 | MATERIALS AND METHODS

2.1 | 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^{10} to 2×2^{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.

2.2 | 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).

2.3 | 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.

2.4 | 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 32×, 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).

2.5 | 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).

2.6 | 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).

2.7 | 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.

2.8 | Clustering of 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 & Schliep, 2019), and ggtree (Yu et al., 2017).

2.9 | Mutation rate

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

2.10 | 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.

3 | RESULTS AND DISCUSSION

3.1 | 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 32× (Supplemental Table S3).

3.2 | 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 32×, 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).

3.3 | 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; L. 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 indicates that the generation of somatic mutations is a stochastic process.

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 & 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.

3.4 | 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 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

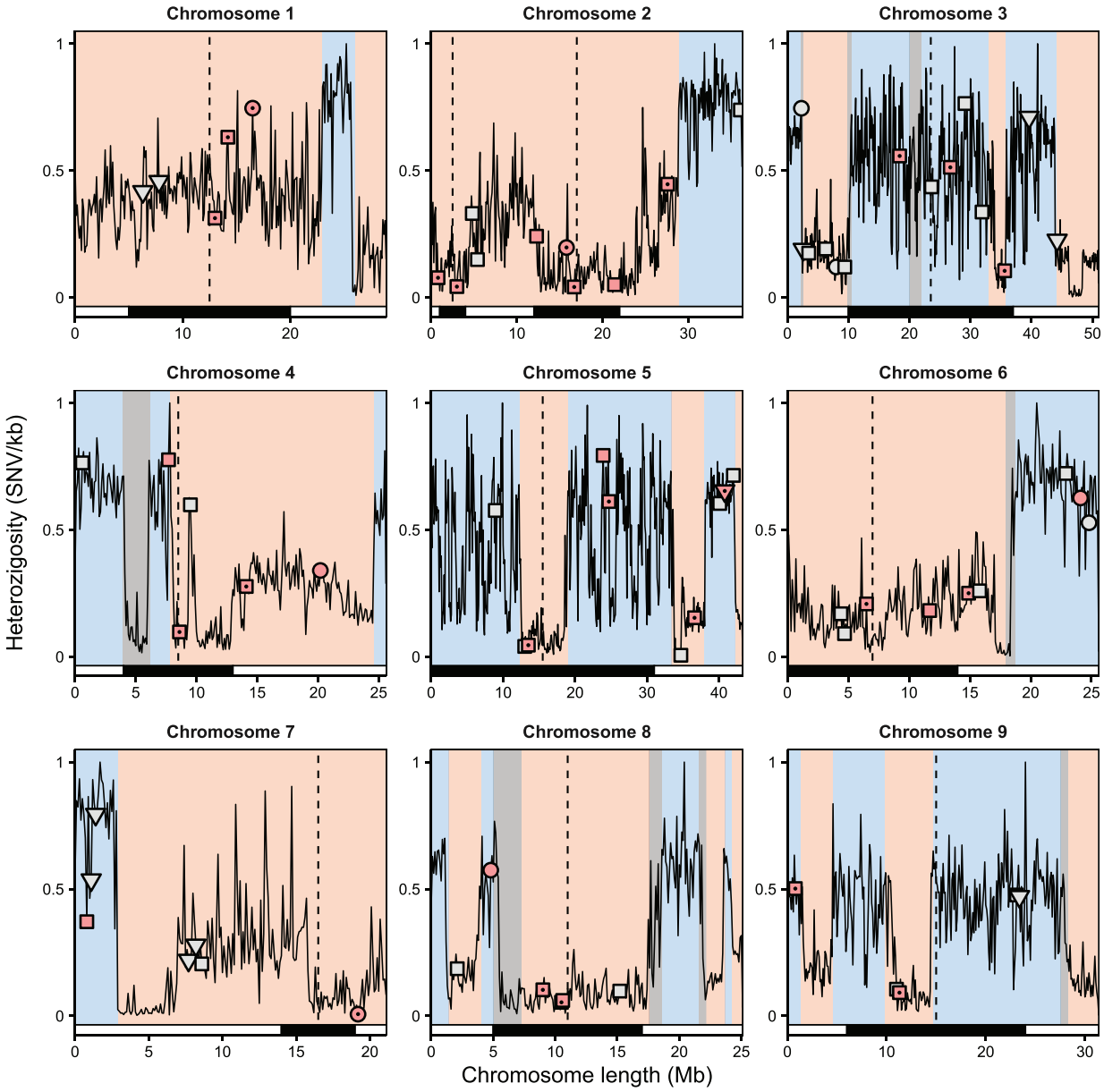

FIGURE 1 Distribution of somatic mutations identified in a 36-yr-old clementine tree along the chromosomes. Heterozygosity is represented in the *y* axis as the nucleotide diversity relative frequency. Pink areas correspond to mandarin–mandarin stretches; blue areas correspond to introgressed regions 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

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

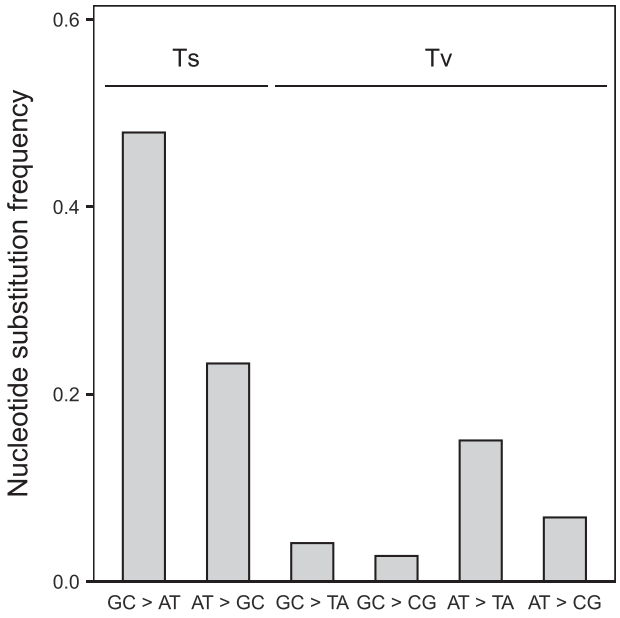

FIGURE 2 Frequency of nucleotide substitution type identified in flushes emerged from the main trunk and branches of different ages in a 36-yr-old clementine tree. Ts, transition; Tv, transversion. Arrow (>) indicates the direction of the nucleotide substitution

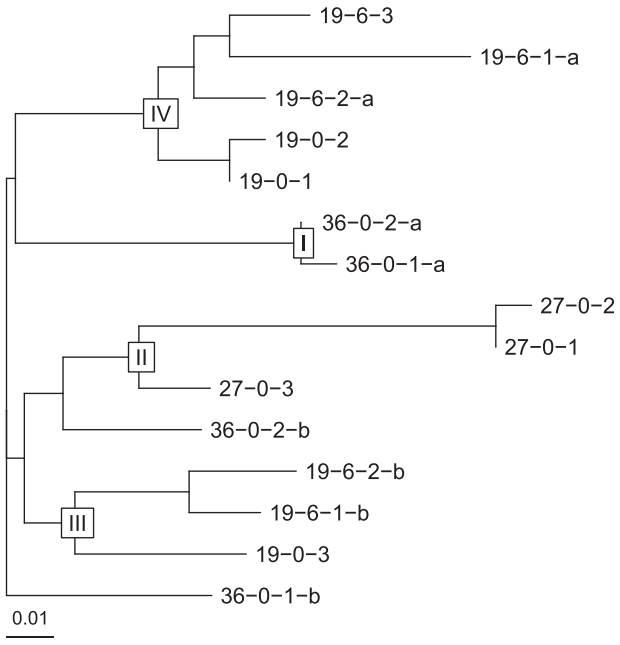

FIGURE 3 Rooted neighbor-joining tree based on the proportion of shared alleles between pairs of samples. Somatic mutations were detected in leaf flushes emerged from the main trunk and branches of different ages in a 36-year-old clementine tree. Roman numerals specify the main four clades revealed by 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

also kept in a dormant bud that sprouted forming the 2-yr-old flush emerged in the basal trunk.

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.

3.5 | Estimation 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 & 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 (L. 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 ($y = 4.0 \times 10^{-9}$, $e^{-0.061x}$, $R^2 = 0.93$, $p = .04$) 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

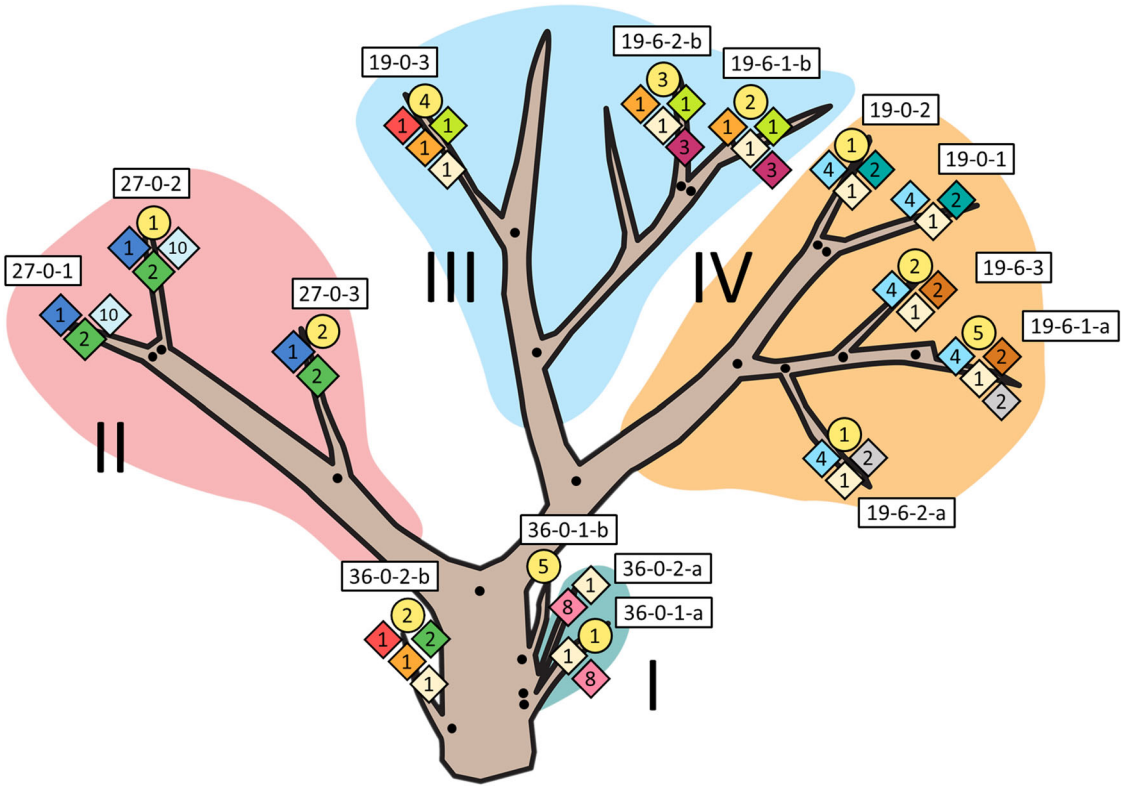

FIGURE 4 Schematic representation of the 36-yr-old clementine tree used for sequencing to study the dynamics of single-nucleotide mosaicism in citrus. Somatic mutations were searched on samples consisting of 1-, 2- and 3-yr-old leaf flushes emerged from 27-, 19-, and 6-yr-old branches and from the most 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

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.

3.6 | 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 (L. 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

TABLE 1 Accumulated somatic mutation rates estimated in a 36-yr-old clementine tree

Samples	Estimated age	Number of nucleotide substitutions	Search space (sites)	Accumulated mutation rates (nucleotide substitutions)		
				Raw	Corrected	Per year
	yr					
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}

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.

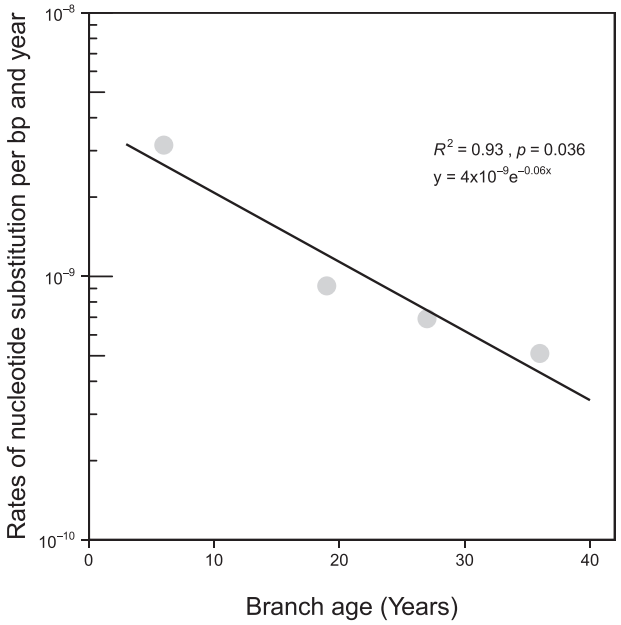

FIGURE 5 Accumulated mutation rate correlation with the age of the branches. Mutation rates were estimated in leaf flushes emerged from 27-, 19-, and 6-yr-old branches and from the most basal part of the trunk of a 36-yr-old clementine tree. All different variants detected in the flushes of each branch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

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.

3.7 | 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 & 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 & Roughgarden, 2012), conferring insect and herbicide resistance (Simberloff & 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.

DATA AVAILABILITY STATEMENT

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

ACKNOWLEDGMENTS

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 IVIA 51915 and 52002. We thank Matilde Sancho, Isabel Sanchis, and Antonio Prieto for laboratory tasks.

AUTHOR CONTRIBUTIONS

Estela Perez-Roman: Data curation; Formal analysis; Investigation; Visualization; Writing-original draft; Writing-review & editing. Carles Borredá: Data curation. Antoni López-García Usach: Resources. Manuel Talon: Conceptualization; Funding acquisition; Investigation; Supervision; Writing-original draft; Writing-review & editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Estela Perez-Roman https://orcid.org/0000-0001-5659-7860
Carles Borredá https://orcid.org/0000-0001-9624-7152
Manuel Talon https://orcid.org/0000-0003-4291-9333

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., … Lippman, Z. B. (2020). Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, *182*, 145–161. 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*, 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*, 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*, 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*, 488–511. https://doi.org/10.1093/genetics/14.5.488

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., 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*, 623–630. https://doi.org/10.1038/nmeth.3407

Folse, H. J., & Roughgarden, J. (2012). Direct benefits of genetic mosaicism and intraorganismal selection: Modeling coevolution between a long-lived tree and a short-lived herbivore. *Evolution*, *66*, 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* (2nd ed.). American Society for Microbiology Press.

Frost, H. B., & Krug, C. A. (1942). Diploid-tetraploid periclinal chimeras as bud variants in citrus. *Genetics*, *27*, 619–634. https://doi.org/10.1093/genetics/27.6.619

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*, 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 and Systematics*, *26*, 423–444. https://doi.org/10.1146/annurev.es.26.110195.002231

Hanlon, V. C. T., Otto, S. P., & Aitken, S. N. (2019). Somatic mutations substantially increase the per-generation mutation rate in the conifer *Picea sitchensis*. *Evolution Letters*, *3*, 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., Ruiz-Rivero, O., Lliso, I., Morillon, R., Tadeo, F. R., & Talon, M. (2007). Physiology of citrus fruiting. *Brazilian Journal of Plant Physiology*, *19*, 333–362.

Ikehata, H., & Ono, T. (2011). The mechanisms of UV mutagenesis. *Journal of Radiation Research*, *52*, 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*, 239–243. https://doi.org/10.1093/molbev/msu302

Klekowski, E. J., & Godfrey, P. J. (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*, 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 [q-bio.GN]. https://arxiv.org/abs/1303.3997v2

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*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lupski, J. R. (2013). Genome mosaicism—One human, multiple genomes. *Science*, *341*, 358–359. https://doi.org/10.1126/science.1239503

Ossowski, S., Schneeberger, K., Lucas-Lledo, 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*, 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*, *35*, 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., … Salse, J. (2018). Oak genome reveals facets of long lifespan. *Nature Plants*, *4*, 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. 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*, 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., … Reymond, P. (2017). Low number of fixed somatic mutations in a long-lived oak tree. *Nature Plants*, *3*, 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*, 49–73. https://doi.org/10.1146/annurev-ecolsys-110218-024955

Simberloff, D., & Leppanen, C. (2019). Plant somatic mutations in nature conferring insect and herbicide resistance. *Pest Management Science*, *75*, 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*, 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*, 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

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*, 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*, 287–292. https://doi.org/10.1007/BF00347587

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag

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*, 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., … Rokhsar, D. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology*, *32*, 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*, 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 of the Royal Society B: Biological Sciences*, *283*, 20161016. https://doi.org/10.1098/rspb.2016.1016

Yu, G., Smith, D. K., Zhu, H., Guan, Y.i, & Lam, T. T.-Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*, 28–36. https://doi.org/10.1111/2041-210X.12628

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Perez-Roman, E., Borredá, C., López-García Usach, A., & Talon, M. (2021). Single-nucleotide mosaicism in citrus: Estimations of somatic mutation rates and total number of variants. *Plant Genome*,e20162. https://doi.org/10.1002/tpg2.20162