Phylogenetic and taxonomic status of Citrus halimii B.C. Stone determined by genotyping complemented by chemical analysis of leaf and fruit rind essential oils

Luro François, Clémentine Baccati, Mathieu Paoli, Elodie Marchi, Gilles Costantino, Marc Gibernau, Patrick Ollitrault, Félix Tomi

Phylogenetic and taxonomic status of *Citrus halimii* B.C. Stone determined by genotyping complemented by chemical analysis of leaf and fruit rind essential oils

François Luro[1*], Clémentine Baccati[2], Mathieu Paoli[2], Elodie Marchi[1], Gilles Costantino[1], Marc Gibernau[2], Patrick Ollitrault[1,3], Félix Tomi[2]

[1]UMR AGAP Institut, Univ. Montpellier, CIRAD, INRAE, Institut Agro, 20230, San Giuliano, France; gilles.costantino@inrae.fr (G.C.); elodie.marchi@inrae.fr (E.M.) francois.luro@inrae.fr (F.L.)

[2]UMR SPE 6134 – Université de Corse – CNRS, Equipe chimie et Biomasse, 20 000, Ajaccio, France; clementine.baccati@gmail.com (CB); gibernau_m@univ-corse.fr (M.G.) ; tomi_f@univ-corse.fr (F.T.)

[3]CIRAD, UMR AGAP, F-20230 Montpellier, France; patrick.ollitrault@cirad.fr (P.O.),

*Corresponding author: Dr François Luro Tel.: +33 04 95 59 59 46; fax: +33 04 95 59 59 37. E-mail address: francois.luro@ inrae.fr (F. Luro)

Abstract

Citrus phylogeny is currently based on genome analysis using molecular markers and sequencing. The 7 pure genetic groups that gave rise to all cultivated citrus underlie the diversity of citrus accessions originating from Asia. However, there are wild citrus forms whose phylogenetic position is unknown, such as mountain citron (*Citrus halimii* B.C. Stone) that was discovered in Malaysia in the early 1970s. We sought to elucidate its status by determining its genetic profile with 30 SSR and InDel markers distributed on the 9 chromosomes of the citrus reference genome as compared to those of the 7 pure genetic groups represented by 4 or 5 varieties each. The genetic study was supplemented by a comparison of the composition of essential oils obtained by fruit peel and leaf hydrodistillation to those of the citrus fruits used for genotyping. The genetic study demonstrated that *C. halimii* is not an interspecific hybrid (low heterozygosity) but rather a true species that shares a common ancestor with kumquats (*Fortunella* sp.), which would have evolved separately. The fruit aromatic profiles confirmed this kumquat/mountain citron relationship but also highlighted the uniqueness of *C. halimii* due to the presence of high proportions of compounds that have never been observed in other citrus fruits, such as germacrene D-8-one (accounting for 8.7% of the leaf essential oil).

Key words: SSR, InDels, allelic diversity, genetic distance, heterozygosity, aromatic compounds, GC-MS, ^{13}C NMR

1 Introduction

Morphological descriptors were widely used before the 1980s in numerical taxonomy studies to elucidate plant genetic diversity and relationships between various species (Ollitrault *et al*., 2020). Based on these descriptors, Barrett and Rhodes (1976) were the first to put forward a hypothesis on *Citrus* phylogeny. They suggested that all cultivated citrus originated from three basic taxa (*C. maxima*, *C. medica*, and *C. reticulata*). Later on, Scora (1988) used essential oil and polyphenol chemical compositions in citrus taxonomic investigations and revealed four true *Citrus* species (*C. halimii*, *C. maxima, C. medica*, and *C. reticulata*). DNA polymorphism techniques, which have been widely implemented since the nineties, contributed to highlighting the phylogenetic structures of citrus (Nicolosi *et al*., 2000; Barkley *et al*., 2006; Garcia-Lor *et al*., 2012; 2013; Curk *et al*., 2016; Shimizu *et al*., 2016). These studies revealed a phylogeny of the *Citrus* genus based on 4 ancestral species: *C. maxima*, *C. reticulata*, *C. medica* and *C. micrantha*. These four species appear to be the ancestors of most cultivated citrus, often with few recombination events, such as sour orange (*C. aurantium*), which is a direct hybrid of pummelo and mandarin, and Tahitian lime (*C. latifolia*), a third-generation hybrid whose genome is an admixture of the four ancestral species (Ahmed *et al.,* 2019).

The first complete reference sequences of the citrus genome were posted in *Phytozome* from 2011 (Wu *et al*., 2014). Since then, genome sequencing has enhanced the phylogenomic profile by providing precise information on the genomes structure of and the meiotic events that generated them (Ollitrault *et al.*, 2020). Phylogenetic hypotheses based on DNA markers, such as SSRs and SNPs, were validated, thereby indicating that many mandarin cultivars were not pure mandarins because they introgressed small parts of the pummelo genome during their evolution (Oueslati *et al.*, 2017; Wu *et al.*, 2018). Recent phylogenomic data confirmed the existence of five pure *Citrus* species: *C. cavaleriei* H. Lev. (including *C. ichangensis* Swingle and *C. latipes* (Swingle) Tanaka), *C. micrantha* Wester, *C. maxima* (Burm.) Merr., *C. medica* L. and *C. reticulata* Blanco (Ollitrault *et al*., 2020). *Fortunella* taxa (kumquats) and *Poncirus* taxa (trifoliate oranges), both originating from northern China (Swingle and Reece, 1967), should now be added to the list of pure species or genetic groups of Asian citrus. In a phylogenetic analysis, Garcia-Lor et al. (2013) observed that *C. reticulata,* also originating from northern China, constituted a single clade with *Poncirus* and *Fortunella.* The different flowering seasons of the three genera could probably explain the differentiation between these three taxa, which evolved in sympatry. A major part of the actual phenotypic diversity of edible citrus should be related to the differentiation between these pure species prior to reticulation and introgression

processes (Ollitrault *et al*., 2020). A close correlation between the genetic and phenotypic diversity was thus observed irrespective of the traits, such as the fruit juice chemical composition in primary metabolites (Luro *et al*., 2011), carotenoids (Fanciullino *et al*., 2005) or leaves and fruit rind aromatic compounds (Liu *et al*., 2013).

The volatile composition of citrus peel oils is generally a mixture of the dominant limonene, other monoterpenes and sesquiterpenes, as well as many oxygenated derivatives (Dugo and Mondello, 2010). Despite this common general profile, each *Citrus* species has a unique organoleptic signature due to a balanced mixture of major constituents and to the presence of minor components such as neral and geranial in lemon (Lota *et al*., 2002) or nootkatone in grapefruit (Paoli *et al*., 2016). Chemotaxonomic analyses based on volatile compounds in both fruit peel and leaves have been shown to be suitable for interspecies phylogenetic studies in various *Citrus* species (Liu *et al*., 2013; Zhang *et al*. 2017; 2019).

Several *Citrus* species could still exist as wild plants or little-altered landraces growing in natural conditions (Bayer *et al.*, 2009). Mountain citron (*C. halimii* B.C. Stone), or so called *limau kadangsa* in Malay language, is one of them (Figure 1).

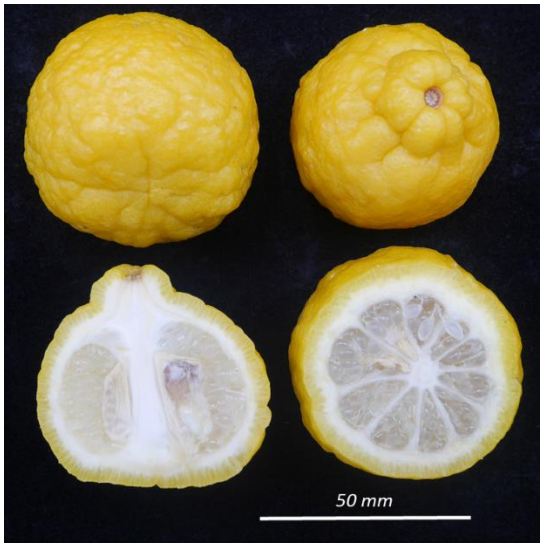

Figure 1: Picture of Mountain citron (*C. halimii*) fruits

It was discovered in Malaysia and Thailand in the 1970s (Stone *et al*., 1973) and its classification and genetic origin has yet to be totally elucidated. Based on morphological and phytochemical data, it was first suggested to be a citron/kumquat hybrid (Scora *et al*., 1976). Stone *et al.* (1973) and Ogawa *et al.* (2001) proposed that *C. halimii* is a natural hybrid between the *Citrus* and *Fortunella* genera, although a RFLP marker analysis did not find any association between *C. halimii* and *Fortunella* (Federici *et al*., 1998). In other studies, *C. halimii* was found to be clustered with *Fortunella* based on isozyme (Herrero

et al. 1996), SSR (Barkley et al. 2006) and AFLP (Pang et al., 2007) marker studies. This relation was supported by cpDNA analysis findings (Bayer *et al*., 2009). Barkley *et al*., 2006 and Bayer *et al.* (2019) rejected the ancestral species status and supported the wild hybrid status because they had not observed any unique alleles in *C. halimii* and concluded that it was an admixture between the kumquat and citron groups, with the majority of its genetic makeup being derived from kumquat. However, Oueslati *et al.* (2017[i]) identified four specific SNPs for *C. halimii* as compared to 78 other accessions of the Aurantioideae subfamily, including *Fortunella* species, based on the sequence of eight plastid genomic regions published by Bayer *et al.* (2019).

C. halimii has seldom been studied from a chemotaxonomical perspective and little information is available on the chemical composition of its tissues, only on leaf cuticle wax (Gulz *et al*., 1987), and its leaf and rind essential oils have only been briefly described (Scora *et al*., 1976). These studies highlighted the highly original chemical features of this citrus fruit, but the comparisons were only made on citrus samples that just included a few cultivated varieties, which were not representative of the genetic diversity of Asian citrus.

Our study aimed to assess the taxonomy and genetic status of *C. halimii* based on codominant DNA molecular markers and the essential oil composition of fruit and leaves. We thus selected varieties from the 7 pure genetic groups representative of the diversity of Asian citrus for comparison with mountain citron.

2 Materials and Methods

2.1. Plant Material

According to recent phylogenomic data (Ollitrault *et al*., 2020) and the Swingle and Reece systematics classification (1967), 35 varieties from 7 pure genetic groups were selected to represent the diversity of Asian citrus (Table 1). Taking into account the phylogenetic relationships described from the genetic data, two genetic groups exist in Papeda, that of *C. cavalereii* which includes *C. ichangensis* and *C. latipes*, and that of *C. hystrix*, which includes *C. micrantha* and *C. macrophylla* (Ollitrault et al. 2020). The accession of Mountain citron (*C. halimii*) was introduced from the *Instituto Valenciano de Investigacion Agrarias* (IVIA) citrus collection (Spain) in 1990, in the form of bud woods that have been grafted to regenerate trees that have been introduced into the INRAE-Cirad citrus collection. Fruit and leaves were randomly picked on trees from the INRAE-CIRAD citrus collection certified as Biological Resource Center (BRC) citrus NF96-600 and located in San Giuliano (France, Corsica): latitude 42°17'N; longitude 9°32'E; Mediterranean climate; average rainfall and temperature 840 mm and 15.2°C per

annum, respectively; soil derived from alluvial deposits and classified as fersiallitic; pH range 6.0–6.6 [31] (Luro *et al.*, 2017).

Table 1: List of citrus varieties used in the analysis of genetic diversity, leaf and fruit peel essential oil composition

Species	Horticultural group	Variety / name	Identity	Analysis		
			ICVN[a]	Genetic	LEO[b]	PEO[c]
Citrus halimii	?	Mountain citron	110302	x	x	x
Citrus medica	Citron	Corsican	100613	x	Lota *et al.* 1999	
Citrus medica	Citron	Diamante	100540	x	Lota *et al.* 1999	
Citrus medica	Citron	Ethrog	100861	x	Lota *et al.* 1999	
Citrus medica	Citron	Humpang	100722	x		
Citrus medica	Citron	Sarcodactylis	100640	x	Lota *et al.* 1999	
Citrus medica	Citron	Poncire	100701	x		
Fortunella hybrid	Kumquat	Fukushu	100325	x	Sutour *et al*. 2016	
Fortunella hindsii	Kumquat	Hong Kong	100743	x	Sutour *et al*. 2016	
Fortunella japonica	Kumquat	Marumi	100482	x	Sutour *et al.* 2016	
Fortunella hybrid	Kumquat	Meiwa	100711	x		
Fortunella marga-rita	Kumquat	Nagami	100490	x	Sutour *et al.* 2016	
Citrus reticulata	Mandarin	Cleopatra	110273	x	Lota *et al.* 2001	
Citrus reticulata	Mandarin	Ladu	100595	x	Lota *et al.* 2001	
Citrus reticulata	Mandarin	Nan feng mi chu	100839	x		
Citrus reticulata	Mandarin	Nanfen Miguan	100700	x	Fanciullino *et al.* 2006	
Citrus reticulata	Mandarin	Sanhu hong chu	100769	x		
Citrus reticulata	Mandarin	Sunki	100705	x	Lota *et al.* 2001	
Citrus maxima	Pummelo	Chandler	100608	x		
Citrus maxima	Pummelo	Eingedi	101130	x	x	x
Citrus maxima	Pummelo	Reinking	100707	x	x	x
Citrus maxima	Pummelo	Deep Red	100611	x	x	x
Citrus maxima	Pummelo	Kao Pan	100321	x	x	x
Citrus maxima	Pummelo	Seedless	100710	x		
Citrus macroptera	Papeda 1	Melanesian	100686	x	Baccati *et al.* 2021	
Citrus micrantha	Papeda 1	Biasong	101115	x	Baccati *et al.* 2021	
Citrus hystrix	Papeda 1	Combava	100630	x	Baccati *et al.* 2021	
Citrus ichangensis	Papeda 2	Ichang 1	100687	x	Baccati *et al.* 2021	
Citrus ichangensis	Papeda 2	Ichang 2	110241	x	Baccati *et al.* 2021	
Citrus ichangensis	Papeda 2	Ichang 3	110240	x	Baccati *et al.* 2021	
Citrus latipes	Papeda 2	Khasi	110243	x	Baccati *et al.* 2021	
Poncirus trifoliata	Trifoliate orange	Rubidoux	110099	x	x	x
Poncirus trifoliata	Trifoliate orange	Pursta	110101	x	x	x
Poncirus trifoliata	Trifoliate orange	Pomeroy	101040	x		
Poncirus trifoliata	Trifoliate orange	Towne	110131	x	x	x

[x]analysis made in the present work; [a]International citrus varietal number; [b]Leaf essential oil, [c]Peal essential oil

2.2 Genotyping

Genomic DNA was extracted from leaf samples using the DNeasy Plant Mini Kit (Qiagen S.A.;) according to the manufacturer's instructions.

The 35 citrus accessions were genotyped with 30 SSR and InDel markers selected according to their distribution on the different genetic linkage groups of the clementine genetic reference map (Ollitrault *et al.*, 2012) and on the reference sequenced genome (Wu *et al*., 2014) (Table 2). PCR was performed as described by Luro *et al*. (2008) in a MWG thermocycler. PCR reactions were performed as simplex experiments in a 6 µl volume with 3 µl of PCR master mix from the Qiagen kit (Type it), 0.2 µL of 10 µM forward primer with a M13 tail at the 5'-end, 0.2 µL of 10 µM reverse primer, 0.2 µL of fluorescently labelled M13-tail (6-FAM, NED, VIC or PET from Applied Biosystems, Foster City, California, USA), 0.12 µL of 5 U/µL *Taq* DNA Polymerase (Taq'Ozyme OZYA001 from Ozyme, Montigny-le-Bretonneux, France) and 10 ng of template DNA. Amplified DNA samples were run on a capillary electrophoresis, based 3130XL genetic analyzer (Applied Biosystems) with an internal standard. Data were analyzed with Genemapper™ software v5.0. Genotyping was performed by the ADNid Company/Qualitech Group (Montpellier, France).

Table 2: Primer sequences, genetic linkage map positions and annealing temperatures in PCR reactions of InDel and SSR markers

Marker	Type	Linkage group position	Forward sequence	Reverse sequence	Ann. temp. PCR (°C)
IDEMA	InDel	1	CTCTTTCTGCTTCCTGACATC	GCCGGTGAATAAAACACAAC	55
Mest121	EST-SSR	1	CAATAATGTTAGGCTGGATGGA	TCCCTATCATCGGCAACTTC	55
Ci02D09	SSR	2	AATGATGAGGGTAAAGATG	ACCCATCACAAAACAGA	55
TAA41	SSR	2	ACATGCAGTGCTATAATGAATG	AGGTCTACATTGGCATTGTC	55
Ci01C07	SSR	2	TTGCTAGCTGCTTTAACTTTA	GTCACTCACTCTCGCTCTTG	55
Mest131	EST-SSR	3	GCTGTCACGTTGGGTGTATG	TACCTCCACGTGTCAAACCA	55
Ci03D12a	SSR	3	CCCACAACCATCACC	GCCATAAGCCCTTTCT	50
Mest256	EST-SSR	3	GAGCAAGTGCGTTGTTGTGT	CATTAAAATATCCGTGCCGC	55
Ci07D06	SSR	4	TCAATTCCTCTAGTGTGTGT	CCTTTTCACAGTTTGCTAT	55
CI01D06a	SSR	4	TTTTTCATCAACAAGACTG	GATCAAAACATTATTCCAA	50
Ci02D04b	SSR	4	AGCAAACCCCACAAC	CTCTCTTTCCCCATTAGA	50
Ci03G05	SSR	4	CCTTGGAGGAGCTTTAC	CCACACAGGCAGACA	50
Mest375	EST-SSR	5	GAAGGAAGAAAAAGAGACCAAAA	CCCCCTTTTGTGATTGTTATG	55
Ci06A12	SSR	5	TTTTTATTTCGGTCTCCTT	CCCAACAAACTCAAACTTC	50
Mest104	EST-SSR	5	TAAAAAGATGGGGCCTTGTG	CCTTATCTTCATCACCTCCGTC	55
Ci01C06	SSR	6	TGGAGACACAAAGAAGAA	GGACCACAACAAAGACAG	50
Mest488	EST-SSR	6	CTTTGCGTGTTTGTGCTGTT	CACGCTCTTGACTTTCTCCC	55
TAA1	SSR	6	AAGAAGAAGAGCCCCCATTAGC	GACAACATCAACAACAGCAAGAGC	55
IDPSY	InDel	6	CCTGTCGACATTCAGGTTAG	CTCATCACATCTTCGGTCTC	55
Ci03B07	SSR	7	TGAGGGACTAAACAGCA	CACCTTTCCCTTCCA	55
Mest107	EST-SSR	7	CCCCATCCTTTCAACTTGTG	GCTGAGATGGGGATGAAAGA	55

Ci01C09	SSR	7	TTGTCCCTTCCCTTTGTA	GACAGAATGGGAGAGGAGA	50
Ci01F04a	SSR	8	TGCTGCTGCTGTTGTTGTTCT	AAGCATTTAGGGAGGGTCACT	55
Mest015	EST-SSR	8	GCCTCGCATTCTCTTGACTC	TTATTACGAAGCGGAGGTGG	55
Ci07B05	SSR	8	CTTTTCTTTCCTAGTTTCCC	TTTGTTCTTTTTGGTCTTTT	50
Ci08C05	SSR	9	CCCTAAAAACCAAGTGACA	TCCACAGATTGCCCATTA	55
IDHYB1	InDel	9	AAAAACAAAGCACCCAGAT	GCCACCAGAACCTGTAATAA	53
Ci07F11	SSR	9	GAAGAAACAAGAAAAAAAAAT	ACTATGATTACTTTGCTTTGAG	50
Ci02B07	SSR	9	TTGGAGAACAGGATGG	CAGCTCAACATGAAAGG	50
Mest149	EST-SSR	9	GGCCATCTTGGTTCAGAGAG	TGCAGCTACCTCGGTAACAC	55

2.3 Analysis of essential oil compositions

2.3.1 Essential oil extraction

For essential oil extraction of mountain citron, pummelos and trifoliate oranges, fruits (100 g of peel used) and leaves (200 g) were randomly picked all around the tree. For other citrus, the essential oil compositions were from the findings of previous studies conducted by our laboratory on kumquats (Sutour *et al*., 2016), mandarins (Lota *et al*., 2001; Fanciullino *et al*., 2006), citrons (Lota *et al*., 1999) and papedas (Baccati *et al*., 2021) sampled from trees of the same citrus collection. The combined use of old and recent data on the composition of essential oils of trees from the same site and grown under the same conditions was possible because the aromatic profiles change very little or not at all over time (Luro et al. 2019).

The fresh materials were subjected to water distillation for 3 h using a Clevenger type apparatus. Peel essential oil (PEO) yields were not calculated because they are influenced by the presence of variable amounts of albedo during epicarp peeling. Distillation yields of leaf essential oils (LEO) were calculated using the essential oil/fresh leaves weight ratio. Each sample was analyzed by dual column gas chromatography and gas chromatography combined with mass spectrometry (GC-MS) in order to determine the chemical composition. To avoid misidentifications, some samples that were selected based on the chromatogram profile were also analyzed via carbon-13 nuclear magnetic resonance (^{13}C NMR) using a method developed in our laboratory (Tomi *et al*., 1995).

2.3.2. Gas chromatography (GC) analysis

GC analyses were performed on a Clarus 500 FID gas chromatograph (PerkinElmer, Courtaboeuf, France) equipped with two fused silica gel capillary columns (50 m x 0.22 mm, film thickness 0.25 µm), BP-1 (polydimethylsiloxane) and BP-20 (polyethylene glycol). The oven temperature was programmed to increase from 60 to 220°C at 2°C/min and then held in an isothermal state at 220°C for 20 min, injector temperature: 250°C; detector temperature: 250°C; carrier gas: hydrogen (1.0 mL/min); and split: 1/60. The relative proportions of the oil constituents were expressed as percentages obtained by peak area normalization without using correcting factors. Retention indices (RIs) were determined

relative to the retention times of a series of n-alkanes with linear interpolation ('Target Compounds' software of PerkinElmer). The essential oil (EO) samples (30 mg) were diluted in 0.5 mL deuterated chloroform ($CDCl_3$).

2.3.3. Mass spectrometry

EOs were analyzed with a PerkinElmer TurboMass detector (quadrupole, Perkin Elmer, Courtaboeuf, France), coupled directly to a PerkinElmer Autosystem XL (PerkinElmer), equipped with a fused silica gel capillary column (50 m x 0.22 mm *i.d*.. film thickness 0.25 µm), and BP-1 (polydimethylsiloxane). Helium was used as carrier gasat 0 .8 mL/min, 1/75 split injection and 0.5 µL was injected. The injector temperature was 250°C. The oven temperature was programmed to increase from 60 to 220°C at 2°C/min and then held in an isothermal state for 20 min. The ion source temperature and energy ionization were set at 250°C and 70 eV, respectively. Electron ionization mass spectra were acquired over a 40–400 Da mass range. Oil samples were diluted in deuterated chloroform with 30 mg of essential oil in 0.5 mL of $CDCl_3$.

2.3.4. NMR analysis

^{13}C NMR analyses were performed on an AVANCE 400 Fourier transform spectrometer (Bruker, Wissembourg, France) operating at 100.623 MHz for ^{13}C, equipped with a 5 mm probe, in $CDCl_3$, with tetramethylsilane (TMS) used as internal reference. ^{13}C NMR spectra were recorded with the following parameters: pulse width (PW): 4 µs (flip angle 45°); acquisition time: 2.73 s for 128 K data table with a spectral width (SW) of 220.000 Hz (220 ppm); CPD mode decoupling; and digital resolution 0.183 Hz/pt. The number of accumulated scans ranged from 2000–3000 per sample (≈ 40 mg of oil in 0.5 mL of $CDCl_3$). Exponential line broadening multiplication (1.0 Hz) of the free induction decay was applied before Fourier transformation.

2.3.5. Identification of individual components

The components were identified via three methods. The first one was a comparison of their GC retention indices (RIs) on polar and apolar columns, determined relative to the retention times of a series of *n*-alkanes with linear interpolation ('Target Compounds' software of PerkinElmer), with those of authentic compounds (McLafferty & Stauffer, 1988). The second one was based on computer matching against commercial mass spectral libraries (McLafferty & Stauffer, 1994; König *et al*., 2001) and by comparison of spectra with literature data (Joulain & König, 1998; Adams, 2007). The last method was a comparison of the signals in the ^{13}C NMR spectra of EOs with those of reference spectra compiled in the laboratory spectral library with the help of laboratory-made software (Tomi *et al*.,

1995; Tomi & Casanova, 2006; Bighelli & Casanova, 2009). In the investigated samples, NMR identified individual components at contents as low as 0.5%.

2.4. Data analyzes

Genetic relationships between the different varieties were analyzed with DARwin v6 software (Perrier *et al.*, 2003) using the weighted neighbor joining method based on the 'Simple matching' similarity index, which took the percentage of common alleles between two citrus samples divided by the total number of observed alleles into account.

Chemical data were analyzed using R v3.6.3 software (2020) with the g-plots packagev3.0.4 to analyze the EO data and determine the relationships between cultivars and components contributing to this diversity (heat maps).

3 Results

3.1 Genetic relationships

The diversity of the 35 citrus trees was found to be organized in 7 genetic groups and 2 isolated genotypes, i.e. *C. halimii* and *C. latipes* (Figure 2). Khasi papeda (*C. latipes*) was connected to the pummelo (*C. maxima*) cluster, while *C. halimii* was linked with the kumquat cluster *(Fortunella* sp.). These phylogenetic relationships between Khasi papeda/pummelo and *C. halimii*/kumquat were quite stable (high bootstrap values of 100 and 68, respectively), but not quite close enough to consider mountain citron as a kumquat and Khasi papeda as a pummelo. Fukushu kumquat was found to be the closest neighbor of *C. halimii*, with a genetic distance of 0.51 between them and an average of 0.59 with the kumquat group. For comparison, the average genetic distance between mountain citron and the mandarin group was 0.79, 0.86 with the pummelo group, 0.90 with trifoliate orange, while the highest distance was 0.92 with the citron group. For Khasi papeda (*C. latipes*) the average genetic distance with pummelo was 0.69.

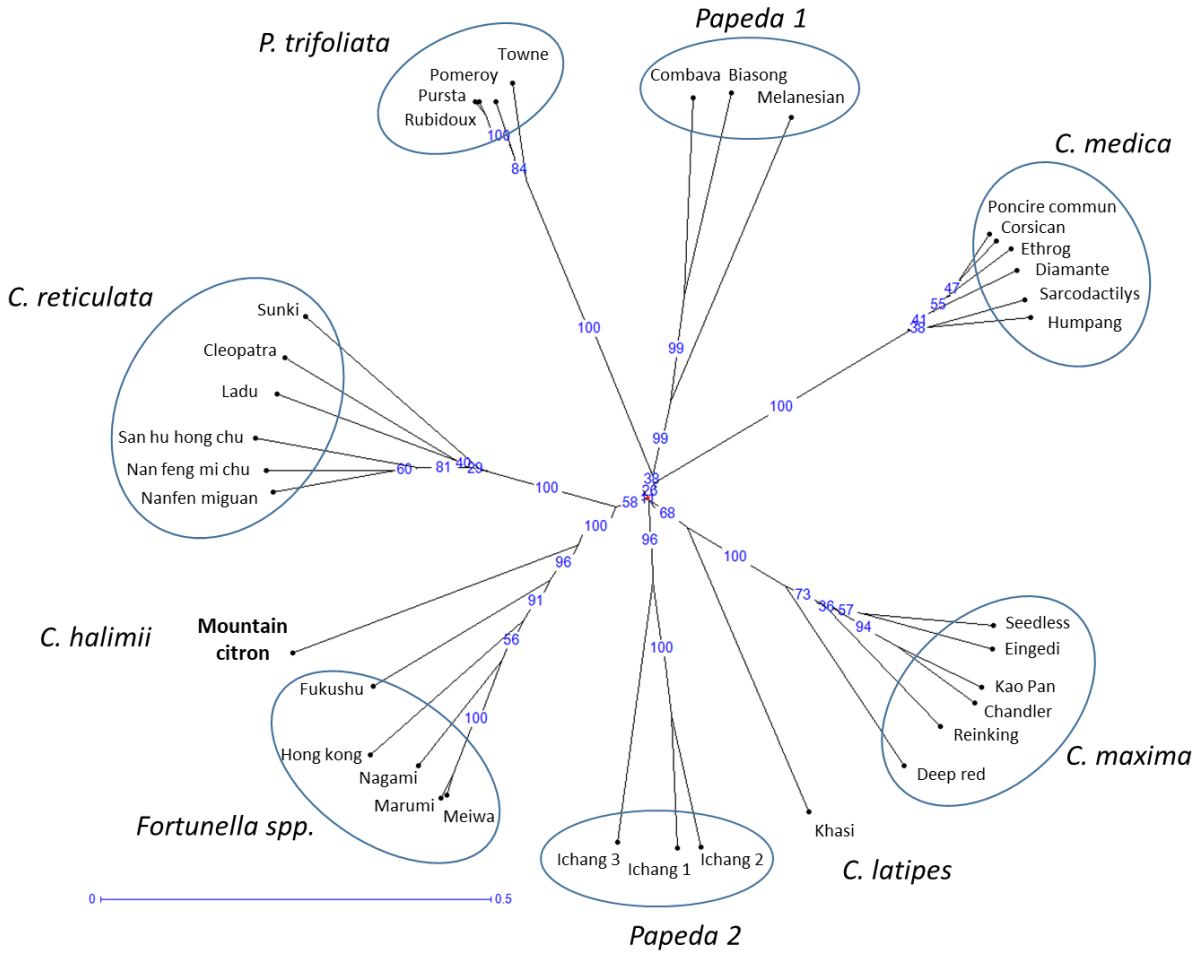

Figure 2: NJ tree of genetic relationships between 35 citrus samples (including *C. halimii*) representing the major citrus genetic groups of Asian origin based on allelic data of 30 nuclear genome markers.

The intervarietal diversity fluctuated markedly depending the genetic group. A high average genetic distance was observed between Melanesian papeda (*C. macroptera*) and the combination *C. hystrix*/*C. micrantha* (0.68), and between Ichang papeda (Ichang 3) and the other two *C. ichangensis* accessions (0.59). The intervarietal average genetic distance reflects the genetic diversity of the species and was found to be particularly high for mandarin (0.46), pummelo (0.40) and kumquat (0.39), while very low for citron (0.23) and trifoliate orange (0.18).

To get a more precise idea of the phylogenetic status of *C. halimii*, some genetic parameters were measured and compared with the different genetic groups identified as clusters in Figure 2 (Table 3). Based on previously reported phylogenetic results, Khasi papeda (*C. latipes*) was excluded from the papeda group 2 (containing the *C. ichangensis* accessions). The proportion of heterozygous loci was very low for citron (8%) and *C. halimii* (13%), relatively high for Khasi papeda (42%) and intermediate (21-32%) for all other genetic groups. *C. halimii* had the lowest allele number per locus but a quite high proportion of specific alleles, i.e. alleles present only in this genotype or group (26%). The

characteristics of *C. halimii* were quite close to those of the *C. medica* group. Khasi papeda had a low number of specific alleles despite its high heterozygosity. A low number of alleles per locus and a very high proportion of specific alleles (45%) distinguished the *P. trifoliata* group. All other genetic groups had quite similar values for the different measured indices, i.e. around 3 alleles per locus and 22–34% specific alleles.

Table 3: Genetic characteristics of the horticultural groups and genotypes (N: genotype number)

	N	% Heterozygous loci	Specific alleles	Allele / locus	% specific alleles
C. halimii	1	13	10	1.27	26
C. medica	6	8	15	1.93	26
Fortunella spp.	5	31	30	3.13	32
C. reticulata	6	25	34	3.33	34
C. maxima	6	21	24	2.93	27
Papeda 1	3	32	34	3.07	37
Papeda 2	3	24	18	2.80	22
C. latipes	1	42	10	1.83	18
P. trifoliata	4	21	23	1.70	45

3.2 Essential oils of C. halimii and other citrus taxa

As usual, fewer compounds were detected in mountain citron PEO (17) than in LEO (37) (Supplemental file: LEO and PEO composition tables). Compounds with a proportion higher than 0.5% in at least in one citrus sample were listed because some of the data were obtained 20 years ago using a less accurate detection method (Lota *et al.*, 1999).

The PEO profile of *C. halimii* was highly dominated by limonene (91.0%), with a noteworthy percentage of *cis*-1,2-limonene oxide (2.1%) and *trans*-1,2-limonene oxide (0.9%), both of which were not found in the other studied species (Figure 3). This oil also exhibited scant quantities of some oxygenated monoterpenes, including *trans*-carveol (0.9%), citronellol (0.6%) and carvone (0.7%). Carvone is another component that was undetected in species other than *C. halimii*. Based on the percentage of limonene, this composition could be compared to those of pummelo (*C. maxima*), kumquat (*Fortunella* sp.) and mandarin (*C. reticulata*), which also exhibited high levels of limonene, i.e. 83.6–93.6%, 93.1–96.3% and 79.0–93.6%, respectively. Other accessions of the different genetic groups exhibited lower limonene levels, with most of the samples containing less than 50% of this compound. Biasong and Combava trifoliate orange varieties exhibited the lowest proportions of limonene (19.5–34.6%). The chemical profile of trifoliate orange varieties differed markedly from that of other citrus accessions by higher average ratios of myrcene (37.5%), α-phellandrene (4.5%), β-phellandrene (11.1%), (E)-β-

ocimene (4.8%) and the presence of specific components such as α-humulene, germacrene B and (2E,6E)-farnesol. This PEO profile differed markedly from that of *C. halimii*. In clustering analyses using compounds with a content of over 1% in at least one citrus sample, *C. halimii* was included in the cluster grouping kumquat, pummelo and mandarin (Figure 3). This association was based mainly on their high limonene contents. The predominance of limonene in PEO clearly influenced the relative percentage of the other compounds and their variations.

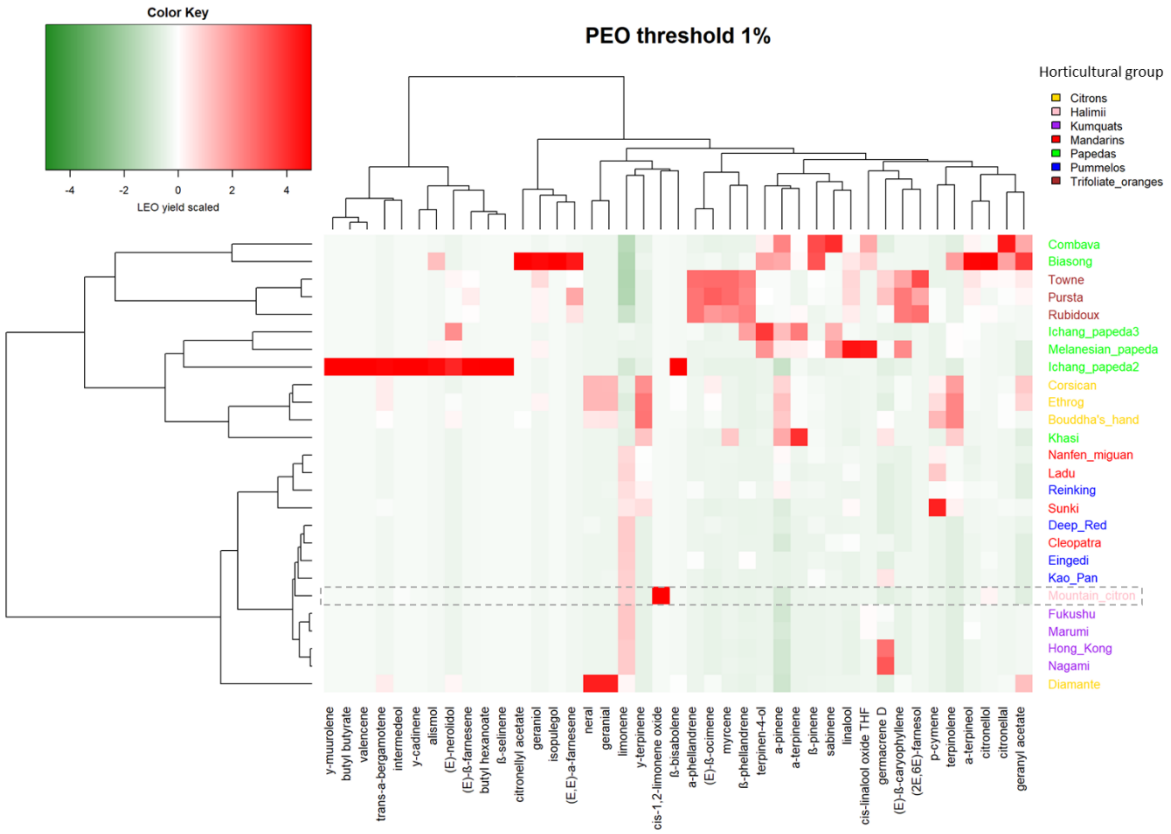

Figure 3: Heatmap of citrus chemical diversity and relationships based on the Jaccard's distance calculated according to the standard proportions of PEO components (>1%), comparing *C. halimii* (highlighted by a dotted box) and 26 other citrus genotypes representing the genetic diversity of Asian citrus forms.

Among citrus LEOs, *C. halimii* exhibited an exceptionally high amount of sesquiterpenes, *i.e.* representing 37.9% of the LEO profile (Supplementary file: Tables of LEO and PEO compositions). In particular, its LEO was characterized by an unusual combination of β-pinene (43.6%), with sesquiterpenes bearing a germacrene skeleton: germacrene D (19.0%), germacrene D-8-one (8.7%) and bicyclogermacrene (2.3%), and an acyclic sesquiterpene: (E)-nerolidol (3.7%).

This documented composition is close to that of pummelo (*C. maxima*) and kumquat (*Fortunella* spp.) for different reasons. On the basis of the monoterpene family, *C. halimii* and pummelo LEOs included very high β-pinene contents, i.e. 23.9–56.7% in pummelo and about 43.6% in *C. halimii*. Contents of this compound only ranged from 0.0 to 4.5% in the rest of the sampling with the exception of three samples, namely Cleopatra mandarin (49.7%), Melanesian papeda (32.4%) and the third *C. ichangensis* accession (44.6%). In addition, pummelo accessions also exhibited sabinene (3.9–9.2%) and limonene (2.4–3.9%) contents similar to that of *C. halimii* (4.8 and 3.3%, respectively). Concerning the sesquiterpene family, the similarity between *C. halimii* and kumquat was highlighted by their high germacrene D content of ≈19.0% for *C. halimii* and 14.9–28.7% in kumquat accessions, while it ranged from 0.0 to 0.6% in the rest of sampling. Kumquat and *C. halimii* also exhibited similar percentages of (*E*)-nerolidol, i.e. 1.0–3.4% and 3.7%, respectively.

This high sesquiterpene content noted in kumquat accessions (82.7–90.8%) and *C. halimii* (37.9%) is very unusual in citrus oils (Figure 4) and did not exceed 13.1% in the other *Citrus* species, while it ranged from 13.8 to 23.6% in the *Poncirus* group.

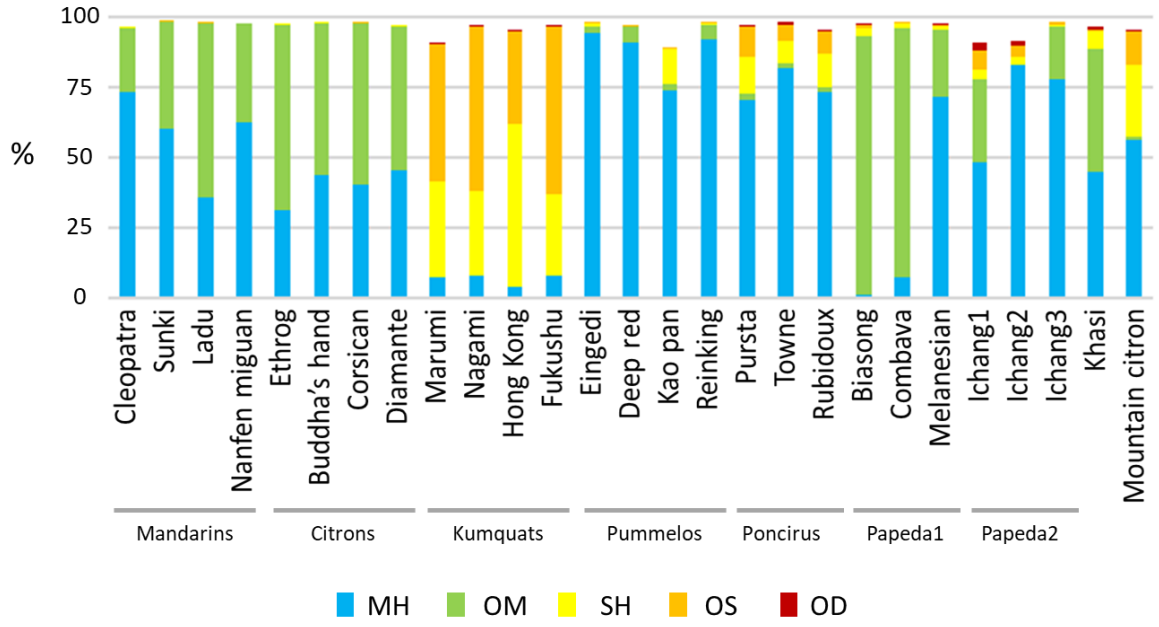

Figure 4: LEO profile characteristics of genetic groups represented by the proportions of the compound families. MH: monoterpene hydrocarbon; OM: oxygenated monoterpene; SH: sesquiterpene hydrocarbon; OS: oxygenated sesquiterpene; OD: oxygenated diterpene

Clustering analysis based on compounds with a content of over 3.5% in at least one citrus sample, revealed greater diversity than that of the PEO composition (Figure 5). Compounds specific to species are observed in LEO. This was the case for kumquat (β-elemol, germacrene D, cis-guai-6-en-10β-ol,

trans-guai-6-en-10β-ol, valerianiol), citron (nerol, neral, geraniol, geranial, limonene, 1,8-cineole, 6-methylhept-5-en-2-one), trifoliate orange (myrcene, α-phellandrene, β-phellandrene, (2E, 6E)-farnesol) and, to a lesser extent, pummelo (β-pinene) and mandarin (γ-terpinene, linalool) where, for each of the latter, one variety was positioned outside the group. The papeda group representatives had very divergent profiles. Note, however, that an exceptionally high proportion of citronellal (≈77%) distinguished Biasong and Combava from other citrus accessions. The fact that the position of *C. halimii* in the *C. maxima* cluster, very far from the kumquat cluster, as well as the low proportion (or absence) of specific compounds of the kumquat group (left part of the heatmap), suggested that the proportion of β-pinene was more of a determining factor for the clustering than the proportion of germacrene D.

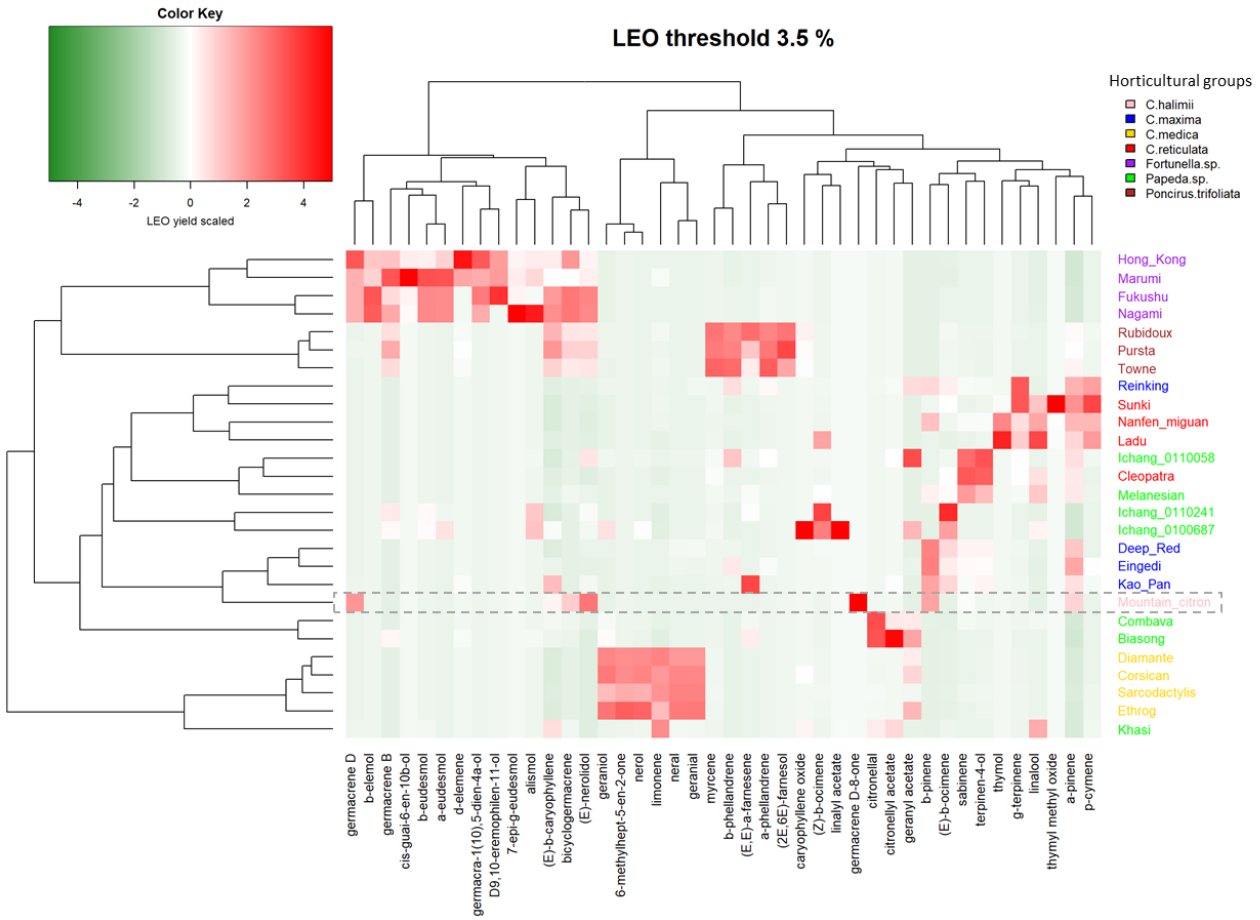

Figure 5: Heatmap of citrus chemical diversity and relationships based on the Jaccard's distance calculated according to the standard proportions of LEO components (>3.5%), comparing *C. halimii* (highlighted by a dotted box) and 26 other citrus genotypes representing the genetic diversity of Asian citrus forms.

In conclusion*, C. halimii* leaf essential oil exhibited a very unique composition, characterized by: i) a low monoterpene/sesquiterpene ratio, *i.e.* 57.4%/37.9%, ii) the presence of germacrene derivatives in appreciable proportions, and iii) the identification of germacrene D-8-one. To our knowledge, this

composition is unique in the *Citrus* genus and germacrene D-8-one may be a chemical marker of *C. halimii* species.

4 Discussion

4.1 Genetic status and origin of C. halimii

C. halimii cannot be classified in the papeda group because of the broad genetic distance between mountain citron and citrus genotypes of the two papeda 1 (*C. micrantha*, *C. hystrix* and *C. macroptera*) and 2 (*C. ichangensis*) genetic groups. Morphological analysis of flowers and leaves already indicated that *C. halimii* did not have the characteristics of this *Citrus* subgenus, as confirmed by our molecular marker study (Stone *et al.* 1973). The relationship with kumquats seems to be real, as previously indicated by Barkley *et al.* (2006) and Bayer *et al.* (2009). However, this relationship is quite distant since the genetic distance between *C. halimii* and kumquats is still greater than 0.5. Bayer et al. (2009) proposed a closer phylogenetic relationship and even suggested that *C. halimii* could be a kumquat hybrid. Our findings refute the idea that the other parent could be a citron, as claimed by Scora *et al.* (1976) and Barkley *et al.* (2006), because the genetic distance between *C. halimii* and *C. medica* is greater than 0.9, thereby reflecting the very low proportion (almost absence) of common alleles between the two species. The genetic distances are also high with respect to the other genetic groups (0.6 to 0.8), which suggests an absence of direct parental relationship between *C. halimii* and *Citrus* species, and even less with *Poncirus trifoliata*. Ten *C. halimii* alleles, representing 26% of the total, were not found in any of the 7 genetic groups assessed in our study. If the hypothesis of hybrid origin were to be put forward, this would imply that another unknown *Citrus* genetic group (or species) would be at the origin of *C. halimii* after a cross with kumquat. Heterozygosity is a very informative index of interspecific hybrid status in citrus. Indeed, secondary species, i.e. the majority of cultivated citrus accessions are interspecific admixtures with interspecific heterozygosity prevailing over a large portion of the genome (Wu *et al*., 2014; 2018). These secondary species have shown high heterozygosity (>0.5) with isozyme (Ollitrault *et al.*, 2003), InDel, SSR (Garcia-Lor *et al.*, 2012; 2013; Curk *et al*., 2016) and SNP markers (Oueslati et al., 2017; Ahmed et al., 2019). Heterozygosity between sweet orange (*C. sinensis*), sour orange (*C. aurantium*), lemon (*C. limon*), and grapefruit (*C. paradisi*) ranged from 0.36 to 0.82 depending on the study and the markers used. As the heterozygosity of *C. halimii* was found to be low (0.13) and equivalent to that of *C. medica* and the high homozygosity of citron was favored by cleistogamic fertilization (Luro *et al*. 2012; Curk *et al*. 2016), it is therefore unlikely that *C. halimii* is an interspecific hybrid. It is hence quite likely that *C. halimii* is a member of a true species with a distant common ancestor with kumquat. Its high homozygosity could be the result of consanguinity due to

small natural population size either with low genetic diversity or, as in citrons, to reproductive biology features leading to selfing.

Apomixis in citrus, which occurs through the development of additional somatic embryos (polyembryony), is an important factor in the fixation of heterozygosity especially in cultivated interspecific hybrids such as sweet orange, grapefruit, sour orange and lemon (Ollitrault *et al.* 2003; Garcia Lor *et al.* 2012, 2013). It is also true that within polyembryonic species such as mandarins (C. reticulata), there are varieties with low heterozygosity such as Cleopatra mandarin or varieties with high heterozygosity such as Willow leaf mandarin (Garcia Lor *et al.* 2015). Pummelos are monoembryonic and their heterozygosity is not very high but not low either because of gametophytic self-incompatibility. The case of citron, which is monoembryonic, is particular because cleistogamy favors self-fertilization and thus the reduction of heterozygosity (Luro *et al.* 2012, Curk *et al.* 2016). *C. halimii* is monoembryonic (Stone *et al.* 1973) but no description of its fertilization mechanism exists. Its heterozygosity is low probably due to self-fertilization but without knowing the causes, which can be due to shift of flowering period compared to other species as for kumquats, cleistogamy as in citrons or restricted population with low level of diversity.

This relatively distant relationship with kumquat is also supported by the morphological characters (Swingle & Reece, 1967; Scora *et al*., 1976): the number of locules is different, i.e. low in kumquats (4-6) and higher in *C. halimii* (7-9); and there is a difference of about 2 months between their respective flowering periods. Other characteristics that differentiate them are the size, shape and texture of the fruit skin and seeds.

4.2 What information can be drawn from the of essential oil compositions?

Previous studies have shown that EO profiles remain stable over time as long as the citrus trees that are the source of the biological material are grown in the same location under the same conditions (Luro *et al*., 2019). Some of the varieties analyzed in this study had been characterized 20 years earlier (Lota *et al.,* 1999; 2001) yet the comparison between aromatic profiles described with a 20 year gap revealed very high stability in the EO composition over time. The EO compositions of *C. maxima* and *P. trifoliata* accessions of the Corsican collection had not been previously analyzed. Within each species, the variety aromatic profiles were close and allowed the detection of compounds specific to these species as well as to the other genetic groups. The trifolate orange profile differed from that described by Scora *et al*. (1969). For example, these authors reported 25.2% of neryl acetate in LEO and 13.5% of γ-terpinene in PEO while in our study the levels of these two molecules in the three *P. trifoliata* varieties were 0 and 0.1%, respectively. The pummelo profiles were similar to that obtained by Zhang *et al.* (2017).

The only description of *C. halimii* LEO found in literature dated back to 1976 and drastically differed from our present description, with limonene (36.5%) and valencene (33.3%) being major components, associated with β-pinene (5.9%), β-caryophyllene (5.6%) and citronellal (4.0%) (Scora *et al.*, 1976). However, the comparison remains difficult because the data was obtained almost 45 years ago and the parameters of analysis were very different (technique of extraction, resolution of the chromatographical column, quality of mass libraries). For instance, and based on the retention indices, we suggest the identification of germacrene D instead of valencene in the LEO profile described by Scora *et al.* (1976).

Concerning PEO, *C. halimii* was clustered with *Fortunella* species and *C. maxima*. For LEO, despite the similar proportions of germacrene D and (*E*)-nerolidol between kumquat and mountain citron, the genetic relationship between *Fortunella* species and *C. halimii* was not found in our cluster analysis. This was probably due to the very low content/absence in of 10 molecules *C. halimii* differentiating it from the *Fortunella* group. The high β-pinene content in LEO seemed to be key for clustering *C. halimii* with pummelos. However, at the compound family level, the relationship between kumquat and mountain citron was clearer with a high proportion of sesquiterpenes.

Several authors have stated that the citrus fruit classification based on volatile compounds is valid at both the genus and species levels (Jing *et al*., 2015; Zhang *et al*., 2017). This is partly justified by the fact that the allopatric evolution of ancestral *Citrus* species resulted in parallel differentiation of the genomes and many phenotypic traits (Ollitrault *et al.*, 2020). The high variability in the results of EO studies published in different parts of the world could probably be explained by environmental impacts on EO expression. The method of aggregation used with aromatic compounds did not reveal the originality of the chemical profile of *C. halimii*, which presents a specific compound in each tissue: germacrene D-8-one in leaves and cis-1-2 limonene oxide in peels. Inconsistently with the genetic analysis, the composition of essential oils revealed a relationship between *C. halimii* and *C. maxima*, only supported by few common metabolites such as α-pinene and β-pinene in the LEO. The genetic markers are independent of environmental and of quantitative effects. The information they give on genetic relationships between species are therefore more robust than analysis of secondary metabolites. The discrepancy between the findings of DNA polymorphism analyses (ours and previous studies) and clustering based on LEO illustrates the limits to the use of these molecules—resulting from complex biosynthesis pathways and under environmental interactions—for phylogenetic classification. In the future, the comparison of whole genome re-sequencing, DNA phylogeny and EO diversity data should help in the assessment of the molecular determinism of EO diversification in citrus.

Overall, the aromatic profile of *C. halimii* is particularly unique, with compounds present that have not been detected in other *Citrus* species. Germacrene D-8-one (Figure 6), or germacra-1(10),4(15),5-trien-8-one, was recently structurally elucidated using 1D and 2D NMR sequences and was reported to be a natural compound of *Isolona dewevrei* (an Ivorian Annonaceae species) essential oil (Kambiré *et al*., 2020). To our knowledge, this component had never been described in leaf citrus oil.

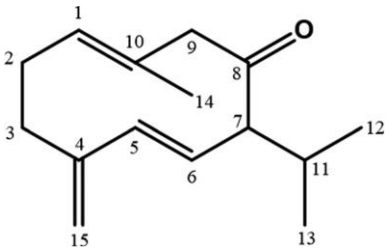

Figure 6: Atomic structure of germacrene D-8-one

5 Conclusion

The findings of this study demonstrated that *C. halimii* is not an interspecific hybrid but is probably a full wild species but with a common ancestor with kumquat. Its uniqueness was noted not only in the specific alleles of SSR or InDel markers but also in leaf and fruit peel volatile profiles that include a unique compound among known citrus fruits. The presence in higher proportions of the sesquiterpene family than in all other *Citrus* species is in agreement with its phylogenetic relationship with *Fortunella spp*, as revealed by DNA markers. This indicates that part of the primitive population of these citrus species migrated northward (to China) from the area of origin, and then evolved into kumquat, and another part of this population which migrated southward (to Thailand and the Malaysian Peninsula) evolved into *C. halimii*, and possibly other as yet unknown related genotypes. The uniqueness of these PEO and LEO volatile compositions, *and the passion fruit aroma of its* fruit skin, make *C. halimii* a special citrus fruit in the Asian citrus group.

Acknowledgements: Thanks to Dr Olivier Pailly responsible of the citrus Biological Resource Center (BRC) for the availability of biological material.

Funding: This work was supported by the European Regional Development Fund under the framework PO FEDER-FSE Corse , France 2014-2020 number 247SAEUFEDER1A, project called Innov'Agrumes (ARR-18/517 CE, synergie number: CO 0009083).

References

Ahmed, D., Comte, A., Curk, F., Costantino, G., Luro, F., Dereeper, A., Mournet, P., Froelicher, Y., Ollitrault, P., 2019. Genotyping by sequencing can reveal the complex mosaic genomes in gene pools resulting from reticulate evolution: a case study in diploid and polyploid citrus. Annals of Botany 123, , 1231-1251.

Adams, R.P., 2007.Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry; Allured Publishing Corporation: Carol Stream, IL, USA, Volume 456.

Baccati, C., Gibernau, M., Paoli, M., Ollitrault, P., Tomi, F., Luro, F., 2021. Chemical variability of peel and leaf essential oils in the *Citrus* subgenus *Papeda* (Swingle) and few relatives. Plants 10, 1117. https://doi.org/10.3390/plants10061117

Barkley, N.A., Roose, M.L., Krueger, R.R., Federici, C.T., 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). Theoretical and Applied Genetics 112, 1519 – 1531.

Barrett, H.C., Rhodes, A.M., 1976. A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. Systematic Botany 1, 105 – 136.

Bayer, R.D., Mabberley, D.J., Morton, C., Miller, C.H., Sharma, I.K., Pfeil, B.E., Rich, S., Hitchcock, R., Sykes, S., 2009. A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. American Journal of Botany 96, 668-685.

Bighelli, A., Casanova, J., 2009. Analytical tools for analyzing Cymbopogon oils. In Essential Oil Bearing Grasses—The Genus Cymbopogon; CRC Press: Boca Raton, FL, USA.

Curk, F., Ollitrault, F., Garcia-Lor, A., Luro, F., Navarro, L., Ollitrault, P., 2016. Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. Annals of Botany https://doi.org/10.1093/aob/mcw005.

Dugo, G., Mondello, L., 2010. Citrus Oils: Composition, Advanced Analytical Techniques, Contaminants, and Biological Activity; Dugo, G., Mondello, L., Eds.; CRC Press: Boca Raton, FL, USA.

Fanciullino, A.L.; Tomi, F., Luro, F., Desjobert, J.M., Casanova, J., 2006. Chemical variability of peel and leaf oils of mandarins. Flavour and Fragance Journal 21, 359.

Fanciullino, A.L., Dhuique-Mayer, C., Luro, F., Casanova, J., Morillon, R., Ollitrault, P., 2005. Relationships between juice carotenoid profiles and genetic diversity within cultivated citrus. Journal of Agricultural and Food Chemistry 54, 4397-4406.

Federici, C.T., Fang, D.Q., Scora, R.W., Roose, M.L., 1998.Phylogenetic relationships within the genus Citrus (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. Theor. Appl. Genet. 96, 812–822.

García-Lor, A., Luro, F**.**, Navarro, L., Ollitrault, P., 2012. Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity; a perspective for genetic association studies. MGG 11, 1-18.

Garcia-Lor, A., Curk, F., Snoussi-Trifa, H., Morillon, R., Ancillo, G., Luro, F., Navarro, L., Ollitrault, P., 2013. A nuclear phylogeny: SNPs, indels and SSRs deliver new insights into the relationships in the "true citrus fruit trees" group (Citrinae, Rutaceae)" and the origin of cultivated species. Annals of Botany 111, 1-19.

Garcia-Lor, A., Luro, F., Ollitrault, P., Navarro, L., 2015. Genetic diversity and population-structure analysis of mandarin germplasm by nuclear, chloroplastic and mitochondrial markers. TGG 11, 123. https://doi.org/10.1007/s11295-015-0951-1

Gulz, P.G., Scora, R.W., Muller, E., Marner, F.J., 1987. Epicuticular waxes of Citrus halimii Stone. *J. Agric. Food Chem.* 36, 716-720.

Herrero, R., Asins, M.J., Pina, J.A., Carbonell, E.A., Navarro, L., 1996. Genetic diversity in the orange subfamily Aurantioideae. II. Genetic relationships among genera and species. Theor. Appl. Genet. 93, 1327–1334.

Jing, L., Lei, Z., Zhang, G., Pilon, A.C., Huhman, D.V., Xie, R., Xi, W., Zhou, Z., Sumner, L.W., 2015. Metabolite profiles of essential oils in citrus peels and their taxonomic implications. Metabolomics 11, 952–963.

Joulain, D., König, W.A., 1998. The Atlas of Spectral Data of Sesquiterpene Hydrocarbons, EB-Verlag: Hamburg, Germany.

Kambiré, D.A., Boti, J.B., Ouattara, Z.A., Thierry, A.Y., Barat, N., Bighelli, A., Tomi, F., 2020. Chemical composition of root and stem bark essential oils from Ivorian Isolona dewevrei: Structural elucidation of a new natural germacrone. Natural Product Research 1-7.

König, W.A., Hochmuth, D.H., Joulain, D. 2001. Terpenoids and Related Constituents of Essential Oils. Library of MassFinder 2.1, Institute of Organic Chemistry: Hamburg, Germany.

Liu, C., Jiang, D., Cheng, Y., Deng, X., Chen, F., Fang, L., Ma, Z., Xu, J., 2013. Chemotaxonomic study of citrus, poncirus and fortunella genotypes based on peel oil volatile compounds deciphering the genetic origin of mangshanyegan (Citrus nobilis Lauriro). PLoS ONE 8, e58411.

Lota, M.-L., de Rocca Serra, D., Tomi, F., Bessiere, J.M., Casanova, J., 1999. Chemical Composition of Peel and Leaf Essential Oils of Citrus Medica, L. and C. Limonimedica Lush. Flavour and Fragrance Journal, 14, 161–166.

Lota, M.-L., de Rocca Serra, D., Tomi, F., Casanova, J., 2001. Chemical variability of peel and leaf oils of 15 species of mandarins. Biochemical Systematics and Ecology 29, 77-104.

Lota, M.L., de Rocca Serra, D., Tomi, F., Jacquemond, C., Casanova, J., 2002. Volatile components of peel and leaf oils of lemon and lime species. J. Agric. Food Chem. 50, 796–805.

Luro, F., Costantino, G., Argout, X., Froelicher, Y., Terol, J., Ollitrault, P., Morillon, R., 2008. Transferability of the EST-SSRs developed on Nules clementine (Citrus clementina Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping BMC Genomics, 9, 287.

Luro, F., Gatto, J., Costantino, G., Pailly, O., 2011. Analysis of genetic diversity in *Citrus.* Pant Genetic Resources 9, 218-221.

Luro, F., Venturini, N., Costantino, G., Paolini, J., Ollitrault, P., Costa, J., 2012. Genetic and Chemical diversity of citron (Citrus medica L.) based on nuclear and cytoplasmic markers and leaf essential oil composition. Phytochemistry, 77, 186–196.

Luro, F., Bloquel, E., Tomu, B., Costantino, G., Tur, I., Riolacci, S., Varamo, F., Ollitrault, P., Froelicher, Y., Curk, F., Pailly, O., 2017. The INRA-CIRAD citrus germplasm collection of San Giuliano, Corsica. In AGRUMED: Archaeology and History of Citrus Fruit in the Mediterranean: Acclimatization, Diversifications, Uses, Fiorentino, G., Zech-Matterne, V., Eds., Collection du Centre Jean Bérard, Publications du Centre Jean Bérard: Naples, Italy, pp. 243–261.

Luro, F., Viglietti, G., Marchi, E., Costantino, G., Scarpa, M.G., Tomi, F., Paoli, M., Curk F., Ollitrault, P., 2019. Genetic, morphological and chemical investigations reveal the genetic origin of Pompia (*C. medica tuberosa* Risso & Poiteau) – an old endemic Sardinian citrus fruit. Phytochemistry 108, 112116.

McLafferty, F.W., Stauffer, D.B., 1989. The Wiley/NBS Registry of Mass Spectral Data, Wiley: New York, NJ, USA, Volume 1.

McLafferty, F.W., Stauffer, D.B., 1994. Registry of Mass Spectral Data, Mass Spectrometry Library Search System Bench-Top/PBM, Version 3.10 d, Wiley-Interscience, Palisade Newfield, NJ, USA.

Nicolosi, E., Deng, Z.N., Gentile, A., La Malfa, S., Continella, G., Tribulato, E., 2000. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theor. Appl. Genet. 100, 1155–1166.

Ogawa, K., Kawasaki, A., Omura, M., Yoshida, T., Ikoma, Y., Yano, M., 2001. 3¢,5¢-Di-C-beta-glucopyranosyl phloretin, a flavonoid characteristic of the genus Fortunella. Phytochem. 57, 737–742.

Ollitraul,t P., Jacquemond, C., Dubois, C., Luro, F., 2003. Citrus. *In*: Hamon, P., Seguin, M., Perrier, X., Glaszmann, X., Genetic diversity of cultivated plants. CIRAD, Montpellier, France, pp. 193-197.

Ollitrault, P., Terol, J., Chen, C., Federici, C.T., Lotfy, S., Hippolyte, I., Ollitrault, F., Bérard, A., Chauveau, A., Cuenca, J., Costantino, G., Kacar, Y., Mu, L., Garcia-Lor, A., Froelicher, Y., Aleza, P., Boland, A., Billot, C., Navarro, L., Luro, F., Roose, M.L., Gmitter, F.G., Talon, M., Brunel, D. A reference genetic map of C. clementina hort. ex Tan., citrus evolution inferences from comparative mapping. BMC Genomics 2012, 13, 593.

Ollitrault, P., Curk, F., Krueger, R., 2020. Citrus taxonomy. In *The Genus Citrus*, Elsevier WP, Talon, M., Caruso, M., Gmitter.J, F.G, Duxford, United Kingdom, pp. 57–81. ISBN 978-0-12-812217.

Oueslati, A., Salhi-Hannachi, A., Luro, F., Vignes, H., Mournet, P., Ollitrault, P., 2017. Genotyping By Sequencing reveals the interspecific C. maxima / C. reticulata admixture along the genomes of modern citrus varieties of mandarins, tangors, tangelos, orangelos and grapefruits. PLOSOne.PONE-D-17-24376R1.

Pang, X.M**.**, Hu C.G., Deng, X.X., 2007. Phylogenetic relationships within *Citrus* and its related genera as inferred from AFLP markers. Genet. Resources Crop Evol. 54, 429–436.

Paoli, M., de Rocca Serra, D., Tomi, F., Luro, F., Bighelli, A., 2016. Chemical composition of the leaf essential oil of grapefruits (*Citrus paradisi* Macf.) in relation with the genetic origin. Journal of Essential Oil Research 28, 265-271.

Perrier, X., Flori, A., Bonnot, F., 2003. Methods for data analysis. In Genetic diversity of cultivated tropical plants, First edition, Hamon, P., Seguin, M., Perrier, X., Glaszmann, J.C., Eds., CIRAD: Montpellier, France, pp. 31−63.

Scora, R. W., Duesch, G., England, A. B., 1969. Essential leaf oils in representatives of the Aurantioideae (Rutaceae). American Journal of Botany 56, 1094-1102.

Scora, R. W., Kumamoto, J., Esen, A., Stone, B.C**.,** 1976. A phytochemical investigation of Citrus halimii. Biochem. Syst. & Ecol. 4, 255-258.

Scora, R.W., 1988. Biochemistry, taxonomy and evolution of modern cultivated citrus. In: Proceedings of the sixth International Citrus Congress. Vol. 1, pp. 277-289.

Shimizu, T., Kitajima, A., Nonaka, K., Yoshioka, T., Ohta, S., Goto, S., Toyoda, A., Fujiyama, A., Mochizuki, T., Nagasaki, H., 2016. Hybrid Origins of Citrus Varieties Inferred from DNA Marker Analysis of Nuclear and Organelle Genomes. PLoS ONE 11, e0166969. https://doi.org/10.1371/journal.pone.0166969.

Stone, B.C., J.B., Lowry, R.W., Scora, K., Jong, 1973. Citrus halimii: A new[ii] species from Malaya and Peninsular Thailand. Biotropica 5, 102 – 110.

Sutour, S., Luro, F., Bradesi, P., Casanova, J., Tomi, F., 2016. Chemical composition of the fruit oils of five Fortunella species grown in the same pedoclimatic conditions in Corsica (France) NPC 11, 259-262.

Swingle, W.T., Reece, P.C., 1967. The Botany of Citrus and its Wild Relatives. Volume 1. In The Citrus Industry, Reuther, W., Webber, H.J., Batchelor, L.D., Eds., University of California: Berkeley, LA, USA, pp. 190–430.

Tomi, F., Casanova, J., 2006. 13C-NMR as a Tool for Identification of Individual Components of Essential Oils from Labiate—A Review. In Proceedings of the I International Symposium on the Labiatae: Advances in Production, Biotechnology and Utilisation, Sanremo, Italy, Volume 723, pp. 185–192.

Tomi, F., Bradesi, P., Bighelli, A., Casanova, J., 1995. Computer-aided identification of individual components of essential oils using Carbon-13 NMR spectroscopy. J. Magn. Reson. Anal 1, 25–34.

Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., et al. 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. 32, 656–662.

Wu, G.A., Terol, J., Ibanez, V., López-García, A., Pérez-Román, E., Borredá, C., Domingo, C., Tadeo, F., Carbonell-Caballero, J., Alonso, R., et al. 2018. Genomics of the origin and evolution of Citrus. Nat. Cell Biol. 554, 311–316.

Zhang, H., Xie, Y., Liu, C., Chen, S., Hu, S., Xie, Z., Deng, X., Xu, J., 2017. Comprehensive comparative analysis of volatile compounds in citrus fruits of different species. Food Chem. 230, 316–326.

Zhang, H., Wen, H., Chen, J., Peng, Z., Shi, M., Chen, M., Yuan, Z., Liu, Y., Zhang, H., Xu, J., 2019. Volatile compounds in fruit peels as novel biomarkers for the identification of four citrus species. Molecules, 24, 4550

---

i